Health Care Financing Grants and Contracts Reports

Economic Analysis of Medical Malpractice: Three Reports

PUBS
RA
1056
.5
D25
1981

Published by the Health Care Financing Administration
Office of Research, Demonstrations, and Statistics

Health Care Financing Grants and Contracts Reports

The Health Care Financing Administration was established to combine health financing and quality assurance programs into a single agency. HCFA is responsible for the Medicare program, Federal participation in the Medicaid program, the Professional Standards Review Organization program, and a variety of other health care quality assurance programs.

The mission of the Health Care Financing Administration is to promote the timely delivery of appropriate, quality health care to its beneficiaries—approximately 47 million of the nation's aged, disabled, and poor. The Agency must also ensure that program beneficiaries are aware of the services for which they are eligible, that those services are accessible and of high quality, and that Agency policies and actions promote efficiency and quality within the total health care delivery system.

HCFA's Office of Research, Demonstrations, and Statistics (ORDS) conducts studies and projects that demonstrate and evaluate optional reimbursement, coverage, eligibility, and management alternatives to the present Federal programs. ORDS also assesses the impact of HCFA programs on health care costs, program expenditures, beneficiary access to services, health care providers, and the health care industry. In addition, ORDS monitors national health care expenditures and prices and provides actuarial analyses on the costs of current HCFA programs as well as the impact of possible legislative or administrative changes in the programs.

ORDS also annually conducts over 200 intramural and extramural research, demonstration, and evaluation projects. The Health Care Financing ***Grants and Contracts Reports*** series presents the final reports from selected ORDS/HCFA-funded extramural projects.

RA
1056.5
.D25
1981
c.2

Health Care Financing Grants and Contracts Report

Economic Analysis of Medical Malpractice: Three Reports

This report was made pursuant to Contract No. 600-76-0150 between the Office of Research, Demonstrations, and Statistics/Health Care Financing Administration and the Rand Corporation.

Patricia Munch Danzon, Project Director
Steven A. Pelovitz, HCFA Project Officer

Published by DHHS
Health Care Financing Administration
Office of Research, Demonstrations,
and Statistics
May 1981

WHY ARE MALPRACTICE PREMIUMS SO HIGH—OR SO LOW?

PREPARED FOR THE HEALTH CARE FINANCING ADMINISTRATION,
U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

PATRICIA MUNCH DANZON

PREFACE

The research reported here was funded by the U.S. Department of Health, Education, and Welfare under Contract No. 600-76-0150, "An Economic Analysis of Medical Malpractice." The work provides empirical evidence on the operation of various components of the malpractice system. Project documentation includes (besides the present report) *Contingent Fees for Personal Injury Litigation* (R-2458-HCFA), *The Disposition of Medical Malpractice Claims* (R-2622-HCFA), and *Determinants of the Frequency of Malpractice Claims* (R-2661-HCFA/ICJ).

This report presents an analysis of the determinants of malpractice premiums reported by physicians in the 1975 and 1976 Surveys of Physician Practice Costs and Income (PPCI), which are nationwide samples of physicians in 15 office-based specialties. The analysis measures the extent to which the wide variation in premiums is attributable to characteristics of the physician, characteristics of the state in which he or she practices, the sponsorship of insurance programs by state medical societies, and the insurance rating bureau.

SUMMARY

Following the medical malpractice "crisis" of 1974-75, the cost of malpractice insurance increased sharply, but not equally for all physicians. In 1975, some physicians paid up to $50,000 and others as little as $75 for coverage. This huge spread bears investigating. Ideally, the tort system of liability provides incentives for physicians to prevent injury. If insurance premiums reflect the individual physician's experience, liability insurance does not interfere with this function. However, if the total cost of claims among physicians is allocated on some other basis, the incentive to prevent injury may be undermined.

This empirical analysis of malpractice insurance premiums examined premiums individual physicians paid in 1974, 1975, and 1976. The sources were two nationwide surveys broadly representing localities and office-based specialties. The study addresses two major issues: (1) how physician characteristics, state characteristics, and coverage differentials explain the differences in premiums among physicians; and (2) how or if institutional factors affect rates. The objective is to measure the variation in premiums and determine the effect on that variation of factors other than loss experience (defined as the history of claims successfully brought against the individual physician).

With a multiplicative system of rating, an individual's premium is determined by multiplying a basic rate by a fixed proportion or differential for each characteristic recognized by the rating structure--specialty, state, and limits of coverage, in the case of medical malpractice insurance. For example, in 1975 and 1976, physicians in the highest risk specialties had rates five times as high as physicians with the same level of coverage in low risk, nonsurgical specialties, regardless of state of residence. Similarly, higher limits of coverage cost all physicians the same multiple of basic limits no matter what their specialty or state residence.

In determining malpractice premiums, specialty is the most important physician characteristic. For this sample of physicians, it accounted for roughly half of the observed variation in premiums. Differences in state characteristics account for a much smaller fraction of the total variation than differences in specialty. About half of the state effect can be attributed to frequency and severity of claims per physician, presence of a discovery rule, and insurance programs sponsored by a medical society. (A discovery rule suspends the running of the statute of limitations until the discovery of the injury.) Premiums are 15 percent higher in states with a discovery rule, after the recent level and trends in frequency and severity of claims are controlled for, indicating the risk in pricing insurance when a discovery rule is in effect.

In many states, medical societies sponsor insurance programs for their members, and it has been charged that these programs limit competition through economies of scale. However, the study finds that on average, such programs do not significantly affect the level of premiums. It has also been alleged that competition is undermined by monopoly power on the seller's side of the market, but the average market share of the leading writer was only 64 percent in 1974. As long as insurance firms compete for the medical society program, having only one successful competitor does not necessarily mean that this firm has market power.

The Insurance Services Office (ISO) develops rates on all lines of property and liability insurance, based on the pooled loss experience of companies that report their losses. Because adherence to ISO rates is purely voluntary, the extent to which they are followed is an empirical question, and the evidence is mixed. The basic multiplicative model used in this study conforms to the rating class system adopted by ISO and fits the data remarkably well. Further, the actual specialty and coverage differentials are similar to those used by ISO. However, available data do not permit determination of how accurately this reflects actual loss experience and would occur even in the absence of ISO. The ISO state differentials add little explanatory power, after actual state loss experience is controlled for, suggesting that ISO state differentials mirror loss experience quite closely and exert little independent influence.

The measured variables--physician and state characteristics--account for about 60 percent of the total variation in premiums in this sample. These characteristics explain much of the variation, confirming the impression from other sources that individual rating experience is not a major factor in determining malpractice premiums. Of course, the study's data refer to 1974-1976, and the malpractice crisis has since wrought many changes. Individual experience rating may affect premiums more today; there are as yet no systematic data to test this conjecture.

For the tort system to deter negligence and create incentives to prevent injuries, physicians' insurance premiums must reflect their individual loss experiences. The rigidity of 1974-1976 premium structures suggests that this is not the case. Policy conclusions should await further research. Data on individual loss experience are needed to evaluate the case for more individual rating. Research is also needed to determine whether this rigid structure is actuarially justified on the basis of interspecialty differences in claims costs and, if so, whether these differences are due to a higher frequency of claims against surgeons rather than to a greater severity of injuries caused by surgeons. If surgical rates are higher primarily because surgical errors are more obvious and easier to prove than other specialties' errors, then the present rating structure may tax surgical specialties disproportionately.

ACKNOWLEDGMENTS

The author would like to thank Christine Peterson for able research and programming assistance. Rand colleagues Robert Bell and William P. Butz provided valuable comment on an early draft. The Insurance Services Office (ISO) kindly provided data on ISO rates. Helen Turin and Joyce Peterson made many constructive editorial comments.

CONTENTS

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS ix

Section

I. INTRODUCTION 1

II. THE MALPRACTICE INSURANCE MARKET 5

III. THE MODEL AND THE DATA BASE 10
  The Model 10
  The Data 12

IV. EMPIRICAL ESTIMATES 16
  Regression Analysis 16
  Analysis of Variance of Premiums ($\log_e$) 23

V. ISO STATE DIFFERENTIALS 26
  Contribution of ISO State Differentials 26
  Determinants of the 1975 ISO State Differentials 28
  Determinants of the ISO Recommended Rate Increase 28

VI. CONCLUSIONS 30

Appendix: SUPPORTING TABLES 35

BIBLIOGRAPHY 45

I. INTRODUCTION

The medical malpractice "crisis" of 1974-75 was characterized by a sharp increase in the cost of malpractice insurance. In a single year, several states had increases of over 300 percent, such that in 1975, some physicians paid up to $50,000 for coverage.[1]

Concern over the cost of malpractice insurance and its contribution to the cost of health care led to the passage of legislation in 49 states.[2] This legislation falls into three categories: (1) changes in the scope of liability and rules of procedure governing medical malpractice claims; (2) measures intended to improve the quality of medical care; and (3) measures intended to ensure the availability and reduce the cost of malpractice insurance. The first two categories--tort reform and quality control--are intended to reduce the cost of insurance indirectly, by reducing the frequency and severity of injuries and of claims. They are discussed in Munch (1980). The third category--insurance related measures--includes: measures to reduce the risk associated with writing this line of insurance (changes in the statute of limitations);[3] authorization for physicians to establish their own insurance arrangements (mutuals or cooperatives); and establishment of residual market mechanisms (joint underwriting authorities--JUAs--or assigned risk pools) to provide coverage for those unable to obtain it in the standard market.

The rules governing tort liability for iatrogenic injury and related insurance arrangements have two major implications. First, they determine who bears the cost of injuries that occur. Second, this assignment of cost affects incentives to engage in risky activities, and thereby affects the allocation of resources to risky activities and the number of injuries that occur. Specifically, how insurance mechanisms allocate the total cost of claims among specialties and among physicians within specialties will affect the allocation of physician manpower among specialties and physicians' incentives to prevent injuries.

---

[1] 1976 Survey of Physician Practice Costs and Income.

[2] NAIC (1977) lists laws passed, by state, through December 1976.

[3] The statute of limitations defines the time allowed for filing a claim after the occurrence of the actionable event.

The ideal tort system of liability for negligence creates incentives to engage in precisely the efficient level of injury prevention.[4] Liability insurance does not interfere with this deterrence function of the tort system if insurance premiums are rated on the basis of individual experience.[5] The structure and basis of insurance rates are therefore crucial elements in the efficient functioning of the malpractice liability system.

This report presents an empirical analysis of the determinants of rates paid for malpractice insurance during 1974 to 1976--the time of the malpractice crisis. In 1975 premiums ranged from $75 to over $50,000, a 600-fold increase. This spread reflects several factors. The level of coverage ranges from $25,000 to $3 million per incident. For any specific level of coverage, the rate depends on the physician's specialty, with higher rates for surgical than for nonsurgical specialties. For a specific level of coverage and specialty, rates vary by locality. One question to be addressed is how much differences in basic rates across states can be attributed to differences in the frequency of claims per physician, the size of awards, and the risk associated with the statute of limitations and the size of the pool of insureds.

A second set of issues relates to the role of institutional factors. In 1974, the structure of the medical malpractice insurance market appears to depart significantly from the competitive ideal of numerous buyers and sellers. There was significant concentration on both the buyer and the seller sides of the market. In many states either a state or a local medical society sponsored an insurance program for its members.[6] Perhaps as a result of medical society programs, the mean market share of the largest writer was 64 percent.[7] When the huge premium increases were filed in 1974-75, this high level of seller concentration was cited as evidence of monopoly power and, in some states, used as a rationale for increased regulation of insurance rates.[8] However, as long as there is competition for the medical society program, having only one successful competitor in any year does not necessarily mean that that firm has market power.[9] Whether rates are higher in states with a high degree of concentration is an empirical question.

---

[4] The efficient level of injury prevention equates the marginal cost of preventing an additional injury with the marginal benefit in terms of damages avoided. For example, see Posner (1972), Brown (1973).

[5] Shavell (1979). With perfect experience rating, the expected cost of damages due to negligent conduct is internalized to the physician through the expected increase in his insurance premium.

Another important institutional factor in the malpractice insurance market is the Insurance Services Office (ISO), which serves as a rating bureau for all lines of property and liability insurance. Its primary function is to develop rates, based on the pooled loss experience of companies that choose to report their loss experience. Membership and adherence to published ISO rates are voluntary, so the potential for ISO to become a cartel is limited. If a firm can perform the functions performed by ISO at lower internal cost, it has little reason to use ISO rates. Because medical societies own the data on their programs, the carriers for these programs have an alternative data source to ISO. It is an empirical question whether ISO in fact exerts much influence over the structure of malpractice rates.

The residual variation in premiums that cannot be explained by the limits of coverage, the physician's specialty, the frequency and severity of claims and the risk in the state of practice provides some indication of the degree to which premiums vary on the basis of the individual physician's loss experience. Because the data base does not contain information on the physician's loss experience, it is not possible to measure experience rating directly.

---

[6] The role of the sponsoring society varies from active involvement in rating and underwriting to merely acting as a billing agent (Kendall and Haldi, 1973). For a comparison of professional society involvement in liability insurance for different professions, see Munch and Smallwood (1979).

[7] Steves and McWhorter (1975).

[8] The competitive rating law was suspended for malpractice insurance in California and New York.

[9] There is evidence of considerable competition for the medical society program in Southern California in the early 1970s (Munch, 1978).

II. THE MALPRACTICE INSURANCE MARKET

The malpractice insurance market has been described in detail elsewhere (Kendall and Haldi, 1973; Steves and McWhorter, 1975). The information provided here is the minimum necessary to understand this study. Because the data are premiums paid during 1974-76, this description pertains to that period and does not reflect subsequent changes.[1]

In 1974, the medical malpractice insurance market was highly concentrated at the state level. Although several thousand firms write property and liability insurance nationwide, fewer than 100 write medical malpractice, and fewer still are writing in any state at any time. In 1974, the market share of the largest writer in a state ranged from 34 to 90 percent, with a mean of 64 percent.[2]

The market was also typically concentrated on the buyer's side, because state or local medical societies sponsored insurance programs. In 1974, the leading carrier in 32 states was writing through a sponsored program, and in four other states the second largest carrier was writing through a program (see Table A.1). The extent of medical society involvement in these programs varies from merely acting as a billing and collection agency to active involvement in underwriting and rate setting decisions.

ISO provides rating bureau services for all lines of property and liability insurance, including medical malpractice. These services include: the pooling of loss statistics of participating companies; devising a classification system; publication of recommended rates for all classes within the classification system; development and publication of policy forms, which define the terms of coverage; filing rates with the Insurance Department in each state; and handling other dealings with regulators.[3] ISO sells its services on a piecemeal basis.

---

[1] These changes include the formation of JUAs, physician-owned mutuals, and cooperatives.

[2] Steves and McWhorter (1975).

[3] All states require filing of insurance rates with an insurance commissioner. The degree of active regulatory intervention in rate setting varies across states and by line of insurance.

Companies are free to purchase any subset of services for any subset of lines or states, or none at all. ISO has no means to enforce, and does not encourage, adherence to its rates. A company has an incentive to use ISO rates to the extent that ISO can make and file rates at lower cost or with greater accuracy, because of a larger data base. There are no comprehensive data on the extent to which companies use ISO rates. Steves and McWhorter (1975) estimate that from 1955 to 1975 between 35 and 45 percent of physicians were insured by companies that reported their loss data to ISO or its predecessors. However, companies may report data but not use ISO rates, and companies may use ISO rates but not report their loss data. Therefore the extent of ISO's influence on the structure and level of rates is an empirical question.

The ISO malpractice coverage is priced in two (or more) tiers. "Basic limits" covers losses up to a specified lower limit per claim and a second higher limit over all claims in the policy year.[4] ISO calculates the cost of coverage beyond these basic limits by applying excess limits factors to the cost of the basic policy.[5] A physician may purchase a third tier of coverage in the form of an umbrella policy, which has a single upper limit and may include coverage for personal liability and premises, in addition to professional liability.

The ISO class structure determines rates on the basis of medical specialty and locality. Medical specialties are grouped into a small number of fairly homogeneous classes. The number of classes distinguished by the ISO rating structure increased from five in 1974 to eight in 1975. For example, the lowest risk class in 1975--physicians, no surgery--includes general and family practitioners, internists, allergists, cardiologists, pediatricians, psychiatrists, and dermatologists. Excess limits factors differ only between two broad categories, physicians and surgeons.[6]

---

[4] ISO basic limits were $5,000 per claim, $15,000 total (5/15) from 1955 to 1975, when they were increased to $25,000 per claim, $75,000 total (25/75).

[5] ISO excess limits factors are shown in Table A.2.

[6] The surgeon differential was reduced to less than 10 percent in 1975. See Table A.2.

The basis of territorial rating is typically the state, but larger states are subdivided. Class differentials and excess limits factors are uniform across territories. The ISO structure implies a uniform ratio of neurosurgeon to pediatrician rates in all states and a uniform ratio of 1,000/3,000 to 25/75 coverage cost.

The only other rating factors recognized by ISO are surcharges for group practice, for nonphysician employees in the practice, and for performance of certain high risk procedures, such as shock therapy or radiation therapy. There is no standard practice of rating on the basis of exposure, either by number of patient contacts or hours worked. Some society programs charge lower rates for part-time physicians.

The factors discussed so far affect rates through their influence on the average or expected claims cost per physician. Standard theory of insurance pricing assumes that expected claims cost is the only factor relevant to competitive rates (other than a loading fee for overhead costs). Risk is assumed to be eliminated because a large number of independent events are pooled.[7] However, in the case of medical malpractice, the conditions necessary to eliminate risk may not hold.

There are at least two reasons why malpractice premiums may incorporate a charge for risk bearing. First, the number of homogeneous policyholders may be too small to eliminate all uncertainty as to mean expected claim cost per policyholder. To the extent the law relating

---

[7] This follows from two basic statistical principles underlying insurance: (1) the variance of the estimate of the mean of a sample of n independent drawings from a population of homogeneous stochastic events, distributed with mean μ and variance σ^2, is equal to the variance of the underlying distribution, divided by the number in the sample, σ^2/n; (2) the variance of the outcome on a portfolio of n components is given by:

$$\sigma_M^2 = \sum_i^n \sigma_i^2 + 2 \sum_{i \neq j} \rho_{ij}\sigma_i\sigma_j .$$

Therefore, in principle, the portfolio may be composed of items with sufficiently negatively correlated outcomes that the variance on the portfolio overall, σ_M^2, is reduced to zero.

to malpractice is state specific, the number of physicians in a state is the maximum potential pool of homogeneous risks.[8]

Second, the outcomes for individual insureds are not independent, to the extent they are similarly affected by changes in social norms for liability and compensation ("societal inflation"). The greater the covariance in expected outcomes for individual insureds, the less the gains from pooling.[9] Covariance is positive in medical malpractice because of the long lag--up to ten years--between the writing of the policy and occurrence of the injury and the payment of claims. The longer the lag, the greater the influence of societal inflation, which affects all insureds alike, on the ultimate claims cost. The risk associated with societal inflation is greater in malpractice than in other lines, such as automobile liability, because of the statute of limitations. Many states have extended the normal term of the statute of limitations for medical malpractice cases, and some have introduced a discovery rule, which suspends the running of the statute until the injury has or "with reasonable diligence should have been" discovered. The cost of insurance is expected to be higher in states

---

[8] If either contiguous states are similar in their legal environments or within a state there is significant heterogeneity in claims experience, then the number of physicians in a state may understate or overstate the size of the potential risk pool.

[9] Assume that for physician i, the expected cost of claims arising out of policy year t is:

$$E(\tilde{C}_{it}) = E[\tilde{A}_t(\tilde{L}_t, \tilde{L}_{t+1}, \ldots, \tilde{L}_{t+10}) + \beta\tilde{X}_{it}],$$

where A_t is an environmental variable common to all physicians;
L_t is a vector of variables determining the sociolegal environment;
X_{it} is a vector of individual characteristics of the ith physician;
a tilde denotes a random variable.
Then the variance around the expected claims cost is:

$$\sigma^2_{C_{it}} = \sigma^2_{A_t} + \beta^2\sigma^2_{X_{it}} + 2\beta \, \mathrm{COV}_{A_t, X_{it}}$$

The larger is $\sigma^2_{A_t}$, the component of variance common to all insureds, relative to $\sigma^2_{X_{it}}$, the less the gains from pooling.

with a discovery rule not only because the frequency of claims is higher but also because the risk is greater.[10]

---

[10] The risk associated with societal inflation has been reduced by measures taken in response to the 1975 insurance crisis. Many states shortened the term of the statute of limitations and of the period allowed for discovery of an injury. In addition, many insurers switched from an occurrence to a claims made form of coverage. An occurrence policy covers all claims arising out of practice in the policy year, no matter how long thereafter the claim is filed. A claims made policy covers claims filed in the policy year, arising out of practice in prior years. Essentially it shifts from the insurer to the insured the risk associated with future trends in the frequency and severity of claims (see Munch, 1978).

III. THE MODEL AND THE DATA BASE

THE MODEL

The model to be estimated is based on the ISO formula for determining rates. Premiums are assumed proportional to expected loss per physician in the state. The factor of proportionality depends on the limits of coverage, physician's specialty and size of practice, and risk and institutional characteristics of the state such as the presence of a discovery rule, market share of the leading writer, and medical society involvement.

$$\begin{aligned} \ln P_{ijk} = {} & \alpha_o + \alpha_1 \ln M_{U_i} + \alpha_2 \ln M_{L_i} + \alpha_3 U_i \\ & + \beta S_j + \gamma_1 \ln \hat{L}_k + \gamma_2 \ln \hat{F}_k + \gamma_3 \hat{X}_k \\ & + \delta_1 Z_i + \delta_2 Z_k + \varepsilon_i \end{aligned} \tag{1}$$

where subscript i denotes physician

subscript j denotes specialty

subscript k denotes state

P = premium

M_U = upper limit of coverage (total over all claims) relative to basic limits

M_L = lower limit of coverage (per claim) relative to basic limits

U = umbrella limit ÷ upper limit.

S = vector of specialty dummy variables

$\hat{L}_k$ = expected basic limits loss per claim in state k

$\hat{F}_k$ = expected frequency of claims per physician in state k

$\hat{X}_k$ = expected ratio of excess to basic limits loss per claim

Z_i = vector of variables describing exposure of physician's practice (number of physicians and nonphysician employees)

Z_k = vector of state characteristics (market share of leading writer, discovery rule, medical society program, number of physicians).

Equation (1) implies that the cost of basic limits coverage is proportional to expected loss per physician in the state of practice. The coefficients α_1, α_2 and α_3 measure the elasticities of premium rates with respect to higher coverage limits. They are therefore a set of continuous proxies for the discrete ISO excess limits factors.

Because claims can be filed several years after the year in which the policy is written (the "policy year") and may take several years between filing and ultimate disposition, the expected frequency and severity of claims are assumed to depend on current levels and recent trends. Following insurance practice, the expected cost per claim is divided into two components, basic and excess limits. The former is the average cost per claim, truncated at whatever the specified basic coverage is ($25,000). The latter is the average cost per claim in excess of the basic limits cut off point. Thus in principle:[1]

$$\hat{L} = L_0 e^{g\tau} \tag{2}$$

and

$$\hat{X} = X_0 e^{m\tau} \tag{3}$$

where $\hat{L}$ = expected basic limits cost per claim for policy year t_0
$\hat{X}$ = expected excess limits cost per claim for policy year t_0
L_0 = basic limits cost per claim paid in t_0
X_0 = excess limits cost per claim paid in t_0
g = expected growth rate in basic limits cost per claim
m = expected growth rate in excess limits cost per claim
τ = average lag from policy year to claim disposition.

The expected frequency of claims against a policy written in year t_0 is the summation of the frequency of claims expected to be filed in each subsequent calendar year, arising out of incidents occurring in the policy year, until the statute of limitations runs out.

---

[1] Equations (2)-(4) are approximations, if τ and T are not constants.

$$\hat{F} = \sum_{i=0}^{T} f_{i0}e^{hi} \tag{4}$$

where $\hat{F}$ = expected frequency of claims against policy year t_0

F_{i0} = frequency of claims filed in ith calendar year after policy year, as of year t_0

h = expected growth in frequency

T = term of statute of limitations.

THE DATA

The data on malpractice insurance premiums are drawn from the 1975 and 1976 surveys of Physician Practice Costs and Income (PPCI) conducted for the Health Care Financing Administration (HCFA).[2] The 1975 sample is confined to office-based physicians in 21 states and five specialties: general practice, general surgery, internal medicine, pediatrics, obstetrics/gynecology. The 1976 sample includes office-based physicians in 15 specialties: allergy, cardiovascular disease, dermatology, gastroenterology, general/family practice, general surgery, internal medicine, neurological surgery, obstetrics/gynecology, ophthalmology, orthopedic surgery, otolaryngology, pediatrics, psychiatry, and urology. All states are represented, but there was intentional oversampling of certain areas and specialties. Physicians working less than 20 hours a week, in prepaid or salaried practice, or in practices of more than 14 in 1975 and 10 in 1976 are excluded from the sample. In the 1975 survey physicians were asked how much they paid for their malpractice insurance and the coverage limits in 1974 and 1975. These are referred to as the 1974 and 1975A samples. In the 1976 survey the same questions referred to malpractice insurance coverage in 1975 and 1976.[3] These are referred to as the 1975B and 1976 samples.

---

[2] For a description of the basic premium data see Greenspan, 1979.

[3] Respondents reporting premiums in excess of $50,000 were excluded on suspicion of reporting error.

The data on the frequency and severity of claims are derived from several sources and do not correspond precisely to the theoretical constructs outlined above.

The ISO *Compilation of General Liability Experience: Professional Liability* reports basic and total limits loss and number of claims incurred, by state, for policy years 1970-74, for companies reporting to ISO. "Incurred loss" is the insurance companies' estimate of what total loss costs will ultimately be for the policy year, evaluated as of March 1975. It therefore includes loss costs on claims already closed, and estimated costs of claims reported but not yet closed and claims not yet reported. The more recent the policy year, the larger the component of conjecture. Basic limits are $25,000 per claim, $75,000 total (25/75). Losses are gross of the cost of disposing of claims (adjustment expense).

The National Association of Insurance Commissioners (NAIC) survey of malpractice claims closed from July 1975 to June 1976 reports the number of claims closed and the total and basic limits loss per claim by state, during this 12 month period, by all carriers who wrote more than $1 million of coverage in any year since 1970. Claims closed during this year relate to incidents that occurred on average 30 months earlier (July 1973). Basic limits are $100,000 per claim. Losses are net of adjustment expense.

The 1970 HEW Survey of claims closed is a stratified random sample of claims closed by all major insurers during 1970. This data base is described in Westat (1973).

The data from these three sources are combined to derive estimates of current levels and growth rates in frequency and severity of claims as follows. The ISO data are used for the current basic and excess limits cost per claim. The NAIC data are used for the current frequency of claims per physician. Growth rates in basic and excess limits are computed as the difference between the NAIC and the ISO basic and excess limits cost per claim. This does not yield pure measures of trends because of the change in the definition of basic limits and treatment of adjustment expense between the two sources. Because these discrepancies apply uniformly across states, comparisons across states should be unaffected.[4] The growth rate of frequency is computed as the difference between the NAIC 1975-1976 frequency and the 1970 HEW frequency of claims closed.[5]

The definition of all variables, means, standard deviations and sources are given in Table 1.

---

[4] To some extent these "growth rates" measure the error in ISO projections. However if the projections as of March 1975 were based on actual paid losses prior to that date, these variables are legitimately interpreted as growth rates.

[5] The ISO data cannot be used for claim frequency because it is not known how many physicians are covered by the companies reporting to ISO.

Table 1

DEFINITION OF VARIABLES, MEANS, STANDARD DEVIATIONS, AND SOURCES

Variable	Definition	Source	Mean (1976)[a]	Standard Deviation (1976)
GP	General practice, allergy, cardiovascular disease	PPCI	.33	
General surgery	General surgery	PPCI	.11	
OBG	Obstetrics/gynecology	PPCI	.12	
Neurosurgery	Neurological surgery	PPCI	.03	
Ophthalmology	Ophthalmology	PPCI	.05	
Orthopedic surgery	Orthopedic surgery	PPCI	.04	
ENT	Otolaryngology	PPCI	.04	
Psychiatry	Psychiatry/child psychiatry	PPCI	.14	
Pediatrics	Pediatrics	PPCI	.25	
Urology	Urology	PPCI	.04	
MDS	Number of physicians (>20 hours per week) in practice	PPCI	1.83	1.50
Other employees	Number of nonphysician personnel	PPCI	4.08	3.50
GP ops.	GP × hours spent on operations and surgical assists. $\log_e$	PPCI	.12	.46
Total limit	Total limit of insurance coverage ÷ 300,000. $\log_e$	PPCI	.78	.98
Claim limit	Limit per claim ÷ 100,000. $\log_e$	PPCI	.92	1.05
Umbrella	Limit of umbrella policy ÷ upper limit; = 1 if umbrella coverage is reported to be zero or is not reported	PPCI	.33	.71
Basic loss	Basic limits incurred loss per claim, policy years 1970-74. $\log_e$	ISO	4.40	.30
Excess ratio	Excess ÷ basic limits incurred loss per claim, policy years 1970-74	ISO	.30	.09
Frequency	Claims closed 1975-76 ÷ number of physicians in patient care, 1974. $\log_e$	NAIC & AMA	-2.64	.37
Basic growth	Growth in basic limits loss $= \frac{\text{Basic limits loss 1975-76}}{\text{Basic limits incurred loss 1970-74}} - 1$	NAIC & ISO	-.14	.37
Frequency growth	Growth in claims per physician $= \frac{\text{Claims closed 1975-76}}{\text{Claims closed 1970}} - 1$	NAIC, HEW & AMA	2.11	1.80
Excess growth	Growth in excess limits ratio $= \frac{\text{Excess/basic limits loss 1975-76}}{\text{Excess/basic limits loss incurred 1970-74}} - 1$	NAIC & ISO	.39	2.19
Pool	Number of physicians in patient care, 1974. $\log_e$	AMA	9.10	1.20
Discovery	Dummy = 1 if discovery rule in 1970	HEW		
Program	Dummy = 1 if largest carrier writing through sponsored program, 1974	Steves & McWhorter	.83	
Program × surgeon	Program × dummy = 1 if surgical specialty		.24	
ISO rate	ISO rate for Class 1 physicians, as of April 1975. $\log_e$	ISO	6.23	.57
ISO pending	Percentage pending increase in ISO rate (+1) ($\log_e$)	ISO	.08	.28
ISO increase	Percentage increase recommended in Class 1 rate, April 1975 (+1) ($\log_e$)	ISO	.86	.41

[a] Means and standard deviations for 1975B are negligibly different from means for 1976.

IV. EMPIRICAL ESTIMATES

REGRESSION ANALYSIS

Appendix Table A.3 reports estimates of the effects of all measured physician and state variables on malpractice premiums, for each of the four samples. The main findings are discussed here, focusing on the 1976 and 1975B samples from the 1976 PPCI survey, which are larger samples and more representative of states and specialties than the 1974 and 1975A samples, from the 1975 PPCI survey.

Specialty Effects

Table 2 reports the specialty differentials, relative to GPs doing no surgery, for nine specialties where estimated differentials were significantly different from zero. The differentials range from +520 percent for neurosurgeons, to -50 percent for psychiatrists. No statistically significant difference was found between GPs and allergists, cardiologists, dermatologists, gastroenterologists and internists. There was a slight increase in the interspecialty spread between 1975 and 1976. The joint hypothesis was tested that there was no difference in rates for general surgery, otolaryngology, and urology; and none between obstetrics/gynecology, neurosurgery, and orthopedic surgery. In 1975 the hypothesis of no difference could not be rejected at the 5 percent level. In 1976 it was rejected because of the increased spread in the highest risk group, with rates for neurosurgeons increasing relative to those for obstetricians and orthopedic surgeons.

Table 2 also reports the class differentials recommended by ISO, under the five class system in operation until 1975 and the eight class system introduced in 1975. Relative to the 1975 ISO differentials, the market substantially compresses the upper tail of the spread, with the markup for the highest risk group (neurosurgeons) being reduced roughly 300 percent. The market also extends the lower end of the distribution, with rates for psychiatrists and pediatricians significantly below those of GPs, although ISO groups these specialties together.

Table 2

SPECIALTY PREMIUM DIFFERENTIALS: ESTIMATED AND ISO

(Relative to general practice no surgery)[a]

	Estimated[b]		ISO	
Specialty	1976	1975	1974	1975
General surgery	3.7	3.8	3.7	5.0
Obstetrics/gynecology	4.2	4.5	6.1	6.0
Neurosurgery	5.2	4.8	6.1	8.0
Ophthalmology	2.1	2.2	1.8	3.0
Orthopedic surgery	5.0	5.3	6.1	8.0
ENT	3.4	3.4	4.9	5.0
Psychiatry	.5	.5	1.0	1.0
Pediatrics	.9	.9	1.0	1.0
Urology	3.4	3.6	4.9	4.0

[a] Includes allergy, cardiovascular disease, internal medicine, gastroenterology, dermatology.

[b] Estimated from Appendix Table A.3, in which differentials are measured relative to pediatricians. These are converted to differentials relative to GP as follows:

$$\text{Reported differential} = e^{\alpha_{spec} - \alpha_{gp}},$$

where α_{spec} = specialty differential relative to pediatricians.

α_{gp} = GP differential relative to pediatricians.

With the data available it is not possible to distinguish the extent to which the observed deviation between market and ISO specialty differentials is due to companies choosing to price independently of ISO rather than simply a lag in adopting the new ISO class structure. The ISO change from a five to an eight class system was filed at different times in different states during 1974-75. Regulatory approval can take several months. Contracts are typically at least a year. The increase in the market spread between 1975 and 1976 is consistent with the hypothesis that the market follows ISO, but with a lag. However, evidence from other sources does show some independence of ISO pricing, at least in medical society programs. For the two programs on which we have data (New York and Arizona) the spread between rates for the highest and lowest risk classes is 5:1, consistent with the regression estimates, and in contrast to the 8:1 ISO spread.[1]

Limits of Coverage and Practice Characteristics

The estimated elasticities of premiums with respect to excess coverage limits per claim (claim limit) and total limits (total limit) are .12 and .10 respectively, in both the 1975B and 1976 samples. Because the excess coverage limits are measured relative to basic limits of 100/300, these estimates can be used to calculate a projected value of the excess limits factor for a 1,000/3,000 policy as follows:

$$P_{1,000/3,000} = 10^{.10} 10^{.12} P_{100/300}$$

$$= 1.66$$

The ISO excess limits factors for 1,000/3,000 relative to 100/300 coverage for surgeons were 1.5 in 1975, 1.64 in 1976.[2] Thus market excess limits factors correspond closely to ISO factors. The estimated elasticity with respect to limits of umbrella coverage is .16 in 1975B, .17 in 1976. ISO does not make rates for umbrella coverage so there is no basis for comparison.

Premiums are between 2 and 3 percent higher for each additional physician or nonphysician employee in the physician's practice. Thus the spillover of liability appears to be small. For general practice and other low risk specialties, premiums are significantly affected by the number of operating hours.[3] This is consistent with the ISO class structure which differentiates between physicians performing no surgery and those performing minor surgery in these specialties. Controlling for operating hours, total number of hours worked per week and weeks worked per year are insignificant. This implies that malpractice insurance is a fixed cost of doing business that is independent of actual exposure. The physician can therefore reduce his average insurance cost per visit by increasing the number of visits. To the extent the rating structure creates incentives to *reduce* the time spent per patient, it appears to be inconsistent with the purpose of deterring negligence.[4]

---

[1] The influence of medical society programs should be partially captured by the two dummy variables, to denote that the leading carrier was writing through a medical society program and an interaction between this and a dummy variable denoting physicians in surgical specialties.

[2] The corresponding figures for physicians are 1.48 and 1.62. See Table A.2. We tested for a difference between physicians and surgeons in the excess limits elasticity and found none.

[3] Estimated elasticities are .24-.26.

State Characteristics

Table 3 reports the effect on premium levels of the characteristics of the claims distribution, measures of risk, and medical society involvement in the state. Standardized coefficients are also reported in order to permit a comparison of the relative effects of different variables.

If standardized coefficients of the loss variables are compared, the characteristic contributing most to the cross state variation in premiums is variation in frequency of claims per physician. Estimates of the elasticity of premiums with respect to claims per physician are .38 in 1975B and .40 in 1976. In the two samples using all 50 states (1976 and 1975B), growth in claims frequency has a small but significant positive effect on premiums (.04 in 1976, .03 in 1975B).[5]

In all samples, incurred basic limits loss per claim is insignificant. The growth in basic limits loss per claim is significantly positive in all samples with standardized coefficients almost as large

---

[4] In other lines of professional liability insurance, premiums are roughly tailored to exposure by rating on the basis of billings. See Munch and Smallwood (1979).

[5] In the other two samples, confined to only 21 states, the significant negative effect of growth in frequency is surprising. This subsample of states is apparently more homogeneous with respect to growth in claims frequency. The coefficient of variation is .51 compared with 1.17 for the full sample (unweighted state correlations), suggesting that the positive effect of growth in frequency for the sample as a whole is contributed by states that experienced extreme changes.

Table 3

EFFECT OF CLAIMS FREQUENCY AND SEVERITY, DISCOVERY RULE, NUMBER OF PHYSICIANS AND PROGRAM ON MALPRACTICE PREMIUMS ($\log_e$)[a]

(Standardized β in parentheses)

Variable	1976 Coefficient (Standardized β)	1975B Coefficient (Standardized β)	1975A[b] Coefficient (Standardized β)	1974[b] Coefficient (Standardized β)	ISO 1975[c] Coefficient (Standardized β)
Basic loss	.086	.049	.237	.100	.918*
	(.024)	(.012)	(.034)	(.014)	(.331)
Basic growth	.339**	.369**	.485**	.473**	.592**
	(.105)	(.108)	(.168)	(.162)	(.331)
Frequency	.384**	.404**	.369**	.531**	.258
	(.130)	(.130)	(.119)	(.171)	(.178)
Frequency growth	.025**	.020**	-.150**	-.167**	.009
	(.041)	(.031)	(-.135)	(-.150)	(.049)
Excess ratio	.413**	.492**	-.068	-.455	.169
	(.035)	(.040)	(-.006)	(-.139)	(.033)
Pool	.020	.031	.193**	.245**	.106
	(.022)	(.033)	(.170)	(.212)	(.194)
Discovery	.194**	.153**	.264**	.112	.107
	(.067)	(.050)	(.081)	(.034)	(.089)
Program	.094**	.094**	.110*	.165**	-.271
	(.042)	(.040)	(.052)	(.077)	(-.228)
Program x surgeon	-.173**	-.122*	-.294**	-.278**	--
	(-.068)	(-.046)	(-.113)	(-.106)	(--)

[a] From Appendix Table A.3.

[b] 21 states only.

[c] Dependent variable is ISO recommended rate for GPs for limits 100/300 as of April 1975, all states excluding Colorado, Illinois, New York, and Texas.

* Significant at the 10 percent level.

** Significant at the 5 percent level.

as those for claim frequency.[6] The ratio of excess to basic limits loss per claim has a significant positive effect, with standardized coefficients of roughly .04 in both samples that include all states.[7] The growth in the excess limits ratio was always insignificant and therefore excluded from the regressions reported here.[8]

Of the two measures of risk, only the discovery rule has a significant effect on premium levels. Premiums are at least 16 percent higher in states with a discovery rule. This is a net effect, after current frequency of claims per physician, which is higher in states with a discovery rule, is controlled for.[9] It is consistent with the hypothesis that states with a discovery rule have greater potential for growth in claims costs, over and above the current level, and greater risk. Therefore, 16 percent is a lower bound estimate of the potential total decrease in premiums obtainable by eliminating a discovery rule.[10]

In the two samples including all states, premiums are unaffected by the number of physicians in a state. There is no evidence of unexploited potential gains from loss pooling in small states, perhaps because of pooling across states with similar legal environments or because nonindependence of risks makes gains from pooling nonexistent.

---

[6] The superior explanatory power of the growth in basic loss costs may be because it is based on paid losses, whereas the level variable (basic loss) is derived from incurred loss data. See Section II.

[7] It is only significant in one of the two smaller samples and the magnitude is very small (-.0006).

[8] In principle, the excess limits ratio should affect only the excess limits factor, not the basic limits premium. However, because we use 100/300 as basic limits, the cost of this level of basic coverage is expected to be affected by the excess ratio which is defined using 25/75 as basic limits.

[9] The gross effect of the discovery rule alone, controlling for number of physicians, but not frequency and severity of claims, is 48 percent.

[10] The potential decrease in premiums does not imply a decrease in total social costs of malpractice, however. Shortening the time for discovery shifts the cost of some injuries from the physician to the victim. To the extent the deterrence to negligence is reduced, the total cost of malpractice is increased.

There is some evidence that accurate prediction of claims cost is limited by the small number of claims. A comparison of the NAIC 1976 paid loss data with the 1970-74 ISO incurred loss data provides a rough measure of the accuracy of predictions. As noted above, the 1970-74 ISO data are projected losses for policy years 1970-74, whereas comparable statistics computed from the NAIC 1976 data yield approximations to the actual outcome for these policy years. A comparison of the two is a measure of the error of the projections. Averaging over state means, predictions are accurate. Predicted total limits loss is \$8,132 and paid total limits loss is \$8,222. However, the coefficient of variation is much smaller for the predicted than the actual values: .25 predicted and .59 actual for total loss. The accuracy of mean predictions but underprediction of variance across states in state means suggests that the sample of claims in each state is too small to yield reliable state specific predictions. This creates a dilemma for rating purposes. Ex ante it may make sense to pool loss data of several states because of no difference in *expected* cost. Ex post, however, there may be large differences in realized costs and therefore the appearance of a cross subsidy.

The estimated effect of a dominant medical society program (Table 3) is to raise the premiums of nonsurgical specialties by roughly 10 percent and to lower the premiums of surgical specialties by 3-8 percent in the 1975B and 1976 samples, which include all 50 states. For the subsample of states represented in the 1974 and 1975A samples, the compressing effect of the program is even greater, with a 17 percent increase for nonsurgical specialties and an 11 percent decrease for surgical specialties.[11] When the interaction is omitted from the regression, the program dummy alone is insignificant. These results imply that medical society programs have no effect on the *average* level of premiums. However, this zero net effect masks the offsetting increase for nonsurgical specialties and decrease for surgical specialties. Further research is necessary on differentials in marginal costs to determine whether this compressing effect of the programs represents a movement toward a more efficient pricing structure or an inefficient cross subsidy. The market share of the leading writer is insignificant in all years. There is no evidence that concentration conveys monopoly power.

---

[11] The larger negative differential for surgical specialties in the 1974 and 1975A samples may be because the only surgical specialties in these samples are obstetrics and general surgery, whereas the 1975B and 1976 samples include lower risk surgical specialties. If the percentage reduction is greatest for highest risk specialties in society programs, the measured effect of a program on all surgical specialties combined is expected to be more negative with the 1974 and 1975A samples.

ANALYSIS OF VARIANCE OF PREMIUMS ($\log_e$)

The percentage of the variance in premiums attributable to physician characteristics and limits of coverage is 55 percent in the two samples (1976 and 1975B) that include a large range of specialties and in which the number of hours spent in operations and surgical assists are controlled for.[12] In both samples 91 percent of the explained variation is attributable to specialty and practice size alone, with liability limits adding only 9 percent after specialty and practice size are controlled for (see Table 4).

The percentage of variance in premiums attributable to variance in the frequency and severity of claims across states is much lower: only 5.4 percent in 1975 and 4.8 percent in 1976 (see Table A.4). The number of physicians in the state and a discovery rule add only .2 percentage points after the loss distribution is controlled for. However, the gross effect of number of physicians and a discovery rule is 2.1 percentage points. Thus there is considerable covariance between these two variables and the frequency and severity of claims.

Table A.5 shows the net contribution of state variables, when physician variables are controlled for. When measured physician characteristics are controlled for, the contribution of measured state variables, including a society sponsored program, is 5.4 percentage points

---

[12] It is 44 percent and 46 percent in the two samples with fewer specialties with no control for operating hours.

Table 4

ANALYSIS OF VARIANCE OF PREMIUMS ($\log_e$): CONTRIBUTION OF SPECIALTY, PRACTICE SIZE, AND LIABILITY LIMITS

Equation Number	Variables	1976 DF	1976 SS	1976 R^2	1975B DF	1975B SS	1975B R^2
1	SP, SZ, LIM	15	1567.6	.552	15	1652.9	.553
2	SP, SZ	12	1425.9	.502	12	1498.2	.501
	(1-2) LIM	3	141.7	.050	3	154.7	.052
Residual		2353	1270.2		2236	1336.6	
Total		2368	2837.8		2251	2989.5	

SP = Nine specialty dummies, GP × number of operating hours.
SZ = MDs, other employees.
LIM = Total limit, claim limit, umbrella.

in the 1975B and 5.7 in the 1976 samples.[13] Of this total contribution of measured state variables, less than 4 percent is attributable to the program. Virtually all of this admittedly small contribution of the program is due to the differentially lower rate for surgical specialties.

In Table A.6 the total explanatory power of physician and state variables is decomposed in the reverse order, to estimate the net contribution of physician characteristics, after measured and unmeasured state characteristics are controlled for. The total explanatory power of measured and unmeasured state characteristics, represented by a vector of dummy variables for each state, is only 10.8 percent in 1975B and 9.3 percent in 1976.[14] With the 1975B and 1976 samples, the estimate of the contribution of physician characteristics, after state is controlled for, is the same (53 percent) as the estimate when state is not controlled for. This implies that with a full sample of states and specialties, the covariance between state and physician effects is zero.

---

[13] Measured state effects add 12.6 percentage points in the 1975A sample and 15 percentage points in the 1974 sample.

[14] It is 28.3 and 25.3 percent in 1974 and 1975A.

This is consistent with the ISO rating structure of uniform specialty differentials across states and a multiplicative model of premiums.

In Table A.7 the total state effect is decomposed into those parts attributable to measured and unmeasured characteristics. First the gross state effect is allowed to differ between surgical and nonsurgical specialties. This increases the total explanatory power of state effects by 1.7 percentage points in 1975B and 1.6 in 1976. The surgery/nonsurgery differential accounts for approximately 14 percent of the total gross state effect. After physician characteristics are controlled for, 45 percent of the total state effect is due to measured characteristics, including the program, and 55 percent to unmeasured characteristics in 1975B. In 1976 total state effect is divided evenly between measured and unmeasured characteristics.[15]

In 1975B and 1976, over 90 percent of the differential state effect for surgeons is due to the program; in 1974 and 1975A the corresponding percentage is 84 percent. To the extent there is a difference in the state effect across specialties, it is apparently largely attributable to the operation of medical society sponsored programs.

---

[15] In 1974 and 1975A measured state effects contribute 64 percent of the total, unmeasured effects only 36 percent.

V. ISO STATE DIFFERENTIALS

The basic building block of the ISO rate structure is the rate for Class 1 physicians, for basic limits coverage, by state. This section estimates the contribution of these ISO state differentials to the observed variance of premiums. The ISO state differentials in 1975 and recommended increases for 1976 are then analyzed.

CONTRIBUTION OF ISO STATE DIFFERENTIALS

Table 5 reports an analysis of variance of premiums in the 1975B and 1976 samples in the 45 states for which ISO rates are available, including as an additional explanatory variable the ISO rate for Class 1 (physicians, no surgery), for 25/75 coverage.[1] If ISO rates are not strongly tied to loss experience and if they are closely adhered to in the market, then addition of the ISO rates to the estimating equation should have a significant increase in the explanatory power. For 1975 the ISO variables are the ISO Class 1 rate as of April 1975 (ISO rate) and the percentage increase in states where an increase was pending (ISO pending). In 1976 the percentage increase recommended by ISO staff (ISO increase) is also included. This is only an approximation for the rate finally filed by ISO and approved by the regulatory authorities. All variables are in natural logarithms.

The results are the same in both years. After physician characteristics and other measured characteristics of the state are controlled for, the ISO variables add 2.2 percentage points to the fraction of variance explained. In both years the 1975 rate and pending increase are highly significant but the recommended increase is not significant in 1976.[2] Although the additional explanatory power of the ISO variables is small in absolute terms, it represents approximately

---

[1] ISO rates are not available for Colorado, Illinois, New York and Texas.

[2] Estimated elasticities of actual rates with respect to ISO rates are .52 in 1975, .54 in 1976. Corresponding standardized coefficients are .34 and .38.

Table 5

CONTRIBUTION OF ISO STATE DIFFERENTIALS

Equation Number	Variables	1975B			1976		
		DF	SS	R^2	DF	SS	R^2
1	X, P, P×S, L, R, I	26	1440.6	.613	27	1366.9	.616
2	X, P, P×S, L, R	24	1390.8	.591	24	1316.4	.594
3	(1-2) I	2	49.8	.022	3	50.5	.022
4	X, P, P×S	17	1245.9	.530	17	1169.4	.527
5	(2-4) L, R	7	144.9	.061	7	147.0	.067
6	P, P×S, L, R, I	9	194.7	.083	10	197.5	.089
7	X, P, P×S, R, I	19	1365.2	.584	20	1282.7	.587
8	(7-4) I, R	2	119.3	.054	3	113.3	.060
Residual		1711	911.6		1794	850.9	
Total		1737	2352.2		1821	2217.8	

X = Specialty, practice size, coverage limits.
L = Loss characteristics (basic loss, frequency, excess ratio, basic growth, frequency growth).
P = Program.
P×S = Program × surgeon.
R = Risk (discovery, pool).
I = ISO (ISO rate, ISO pending, ISO increase).

25 percent of the total explanatory power of measured state variables, which is only 8.3 and 8.9 percentage points in these two samples.

The analysis above measures the marginal contribution of ISO state differentials after controlling for parameters of the loss distribution. An alternative approach is to compare the explanatory power of ISO rates on the one hand and parameters of the loss distribution on the other. This comparison yields the conclusion that ISO rates and parameters of the loss distribution are very close substitutes, with the ISO rates conveying slightly less information. Adding the parameters of the loss distribution alone adds 6.1 percentage points in 1975B and 6.7 in 1976. The corresponding figures if the ISO variables are substituted for the parameters of the loss distribution are 5.4 and 6.0.

DETERMINANTS OF THE 1975 ISO STATE DIFFERENTIALS

The finding that ISO state differentials and the parameters of the loss distribution are almost perfect substitutes in explaining variation in malpractice premiums raises the obvious question of how far the ISO rates can themselves be explained by the parameters of the loss distribution and the other state variables. Table A.8 shows that the loss variables, the discovery rule, the medical society program dummies, and the number of physicians in the state explain 38.7 percent of the variation in ISO rates. Of this, the net contribution of operation of a medical society program is 4.3 percentage points or 11 percent of the total. The coefficient on the program dummy is -.27 (significance level 11 percent). This implies that concentration on the buyer side of the market lowers the rates recommended by ISO.

Table 3 reports the actual and standardized coefficients of the loss variables from the ISO rate equations. When these coefficients are compared with those estimated from the 1975B market premium equations, the major difference is that ISO rates are more sensitive to the frequency in claims per physician.[3]

DETERMINANTS OF THE ISO RECOMMENDED RATE INCREASE

In April 1975 ISO recommended rate increases that averaged 180 percent and ranged from 10 percent to 500 percent. These proposed increases were based on expected loss ratios at current premium levels.[4] Table A.9 reports regressions of these proposed increases on the prevailing level of rates, the pending increase, if any; the expected loss ratio at current premium levels; and the measures of the growth rates in basic limits loss, claims per physician, and ratio of excess limits loss to basic limits loss. Most of the explanatory power comes from the current level of rates (negative relation) and a pending increase (positive relation). Although all the growth variables are positive, significance levels are low, both individually and in the aggregate. Coefficients of the growth variables underestimate the influence of growth in claims costs on the recommended rate increase to the extent the expected loss ratio incorporates expected trends. However, when included alone the expected loss ratio is not highly significant, has less explanatory power than the vector of growth variables alone, and adds only an additional 3.1 percentage points.

---

[3] t-statistics are generally lower in the ISO equation, which is based on only 45 observations.

[4] The loss ratio is the ratio of loss costs to premiums.

These findings suggest that the large recommended premium increases occurred in states where existing rates were low in absolute terms, rather than relative to the expected level of claims, based on recent trends in frequency and severity. Other evidence presented here has shown that ISO rates before the big increases reflected state loss characteristics quite closely. The proposed 1975 increases appear to be an exception. It is tempting to attribute to this the apparent failure of the 1976 ISO proposed increase to influence market rates in 1976. However, the data used so far on state specific claims costs and trends are too crude for this conclusion to be more than a hypothesis worth exploring further.

VI. CONCLUSIONS

It was argued initially that if the tort system is to create incentives for an efficient allocation of resources to injury prevention, liability insurance premiums must be tailored to the expected losses of the individual physician. This is the ideal in a world in which the insurer can costlessly observe the quality of care exercised by each physician and hence his expected loss costs. In practice, the insurer can obtain more accurate information about expected loss costs only at increasing cost. The optimal degree of experience rating therefore involves a tradeoff between the costs and the benefits of rating premiums more accurately.

Because data on individual physician experience are not available, this study has not attempted to define the optimal degree of experience rating or to evaluate the efficiency of existing rating practices. The more limited objective has been to measure the variation in premiums paid by a nationwide sample of physicians and to determine how far that variation can be attributed to factors other than individual loss experience. Specifically, we have measured the effect on premiums of observed characteristics of the physician (specialty, practice size, hours worked), the limits of coverage, and the characteristics of the state in which he practices (expected loss per physician, risk, etc.).

A second issue has been to measure the effect on premiums of institutional factors that have been alleged to impede competition in the medical malpractice insurance market--in particular the medical society sponsored programs, the high level of seller concentration, and the advisory role of the Insurance Services Office.

The data on premiums paid in 1975-76 by a nationwide sample of physicians show a spread of over 600 percent, from \$75 to over \$50,000 per year.[1] This huge range is well explained by a multiplicative system of rating. State differentials for the lowest risk class of physicians (GPs, no surgery) for basic limits of coverage are multiplied by specialty differentials and excess limits factors. Any given level of coverage costs a surgeon roughly five times as much as it would cost a general practitioner, regardless of the state of residence; and similarly, higher limits of coverage cost all physicians the same multiple of basic limits, no matter what their specialty or state of residence.

---

[1] This is not a random sample of physicians but is broadly representative of localities and office-based medical specialties.

Specialty is the single most important characteristic of the physician in determining his malpractice premium. Roughly 50 percent of the observed variation in premiums (measured in natural logs) for this sample of physicians can be "explained" by specialty alone. High risk surgical specialties (neurosurgery, orthopedic surgery) pay five times as much as low risk nonsurgical specialties (general practice, internal medicine, etc.). The specialty premium differentials in 1975 and 1976 are very close to those ISO recommended before it switched from a five class to an eight class system, starting in 1974. It is not possible to determine whether the more compressed rate structure observed in the data reflects independent rating decisions by individual insurers or is simply due to a lag in adopting the new ISO eight class structure because of regulatory delay and long term contracts. Whatever the reasons for the specialty premium differentials, their uniformity across states is striking.

Premiums are apparently unrelated to the number of hours worked, which is a rough measure of exposure. This implies that malpractice insurance is a fixed cost of doing business, such that the cost *per patient contact* falls as the number of patients seen increases. If the purpose of the tort system is to increase incentives to take care, this is a perverse aspect of the rating system.[2] The estimated elasticities of premium cost with respect to coverage limit per claim and in total are 0.12 and 0.10 respectively. Differences in limits of coverage account for 5 percent of the variation in premiums in 1975 and 1976.

---

[2] Except to the extent that the physician's skill increases with the number of patients seen.

State characteristics account for some 11 percent of the variance in premiums. Roughly half of the state effect can be accounted for by difference in frequency and severity of claims per physician, risk, and the influence of medical society programs. The remainder must be attributed to unmeasured factors. Of the two measures of risk, the presence of a discovery rule appears to be more important than the number of physicians in the state (a proxy for the limits on pooling). Premiums were approximately 15 percent higher in states with a discovery rule. Because this effect is measured net of the current levels and recent trends in frequency and severity of claims, it may be attributed to the greater risk in pricing insurance in the presence of a discovery rule. Although the number of physicians in the state has no measurable effect on the level of premiums, the discrepancy between the variance of projected and realized loss costs across states suggests that the small number of insureds does contribute to the riskiness of writing malpractice insurance and therefore provides some justification for pooling state data for rating purposes.

Medical society programs are associated with lower premiums for surgical specialties but higher premiums for nonsurgical specialties. These effects cancel out, so that when premiums are averaged over all specialties, they are neither higher nor lower in states where the leading carrier writes through a sponsored program. There is no evidence that programs lower premiums because of economies of scale or monopsony power.[3] Similarly, there is no evidence that rates are affected by concentration on the seller side, as measured by the fraction of the market written by the largest firm.

Conclusions as to ISO's influence are tentative. The basic multiplicative model corresponds to that used by ISO. The estimated elasticities of cost of higher levels of coverage correspond closely to the excess limits factors recommended by ISO. The significant explanatory power of the specialty differentials implies uniformity of specialty differentials across states, which is consistent with the ISO class structure. The estimated specialty differentials are similar but not identical to those recommended by ISO, but it is not possible to determine the extent to which the discrepancies are due to independent pricing, influence of medical societies, or merely lag in adopting the new ISO differentials.

---

[3] The only evidence against the conclusion that sponsored programs do not affect the average level of premiums is that rates recommended by ISO were lower in states with a sponsored program. However, the significance level was low (11 percent). Moreover, ISO rates are probably derived from loss data from less representative subsamples in states where the majority of the market is written through a sponsored program. If so, these rates do not indicate what rates would be in the absence of a sponsored program.

The ISO state differentials add very little explanatory power, after other measured characteristics of the state are controlled for (frequency and severity of claims, risk, etc.). In fact, the recommended ISO state differentials and the characteristics of the loss distribution are very close substitutes in terms of explanatory power. At least for the period before the large increases, there is no evidence that ISO exerted an independent influence on the relative cost of coverage in different states. Whether ISO influenced the absolute level of premiums is beyond the scope of this study to determine.

Without more detailed data on claims costs, it is not possible to derive normative conclusions about the efficient rating structure and efficient degree of individual experience rating. Similarly, it is not possible to determine how closely the level and degree of uniformity of excess coverage and specialty differentials conform to loss experience. However, the actual structure of premiums is apparently quite rigid. The high degree of explanatory power of the simple estimating equation that incorporates only specialty, limits, and state differentials suggests that the amount of variation due to individual experience is small. The residual unexplained 40 percent is an upper bound, because it includes variance due to reporting error, different types of coverage (claims made rather than occurrence), and all other unmeasured variables in addition to individual experience rating.[4]

---

[4] If the individual rating takes the form of exclusions of coverage or deductibles, the same premium level would in fact represent different prices, so the unexplained 40 percent of the variance in premiums is not an upper bound on individual rating.

Policy conclusions must await further research with better data. Data on individual loss experience is necessary to evaluate the case for more individual rating. A related question for future research is whether the large and apparently rigid specialty differentials are in fact actuarially justified on the basis of interspecialty differences in claims costs and, if so, the extent to which these differences are due to a higher frequency of claims against surgeons rather than a greater severity of injuries caused by surgeons. It seems implausible that surgeons are more negligent than nonsurgical specialists. If the surgical premium rates are higher because surgical errors are more costly, then it is optimal that premiums be higher to induce a higher level of care. However, if surgical rates are higher simply because surgical errors are more obvious and easier to prove, then the current specialty rating structure may impose a disproportionately high tax on surgical specialties.

Appendix

SUPPORTING TABLES

Table A.1

MARKET STATUS: YEAR END, 1974

State	Principal Carrier(s)	Percentage of Exposures	Principal Competitor
Alabama	Employers of Wausau[a]	60	St. Paul
Alaska	Lloyd's of London	N.A.	N.A.
Arizona	Travelers	50	Signal/Imperial
Arkansas	St. Paul Fire & Marine	80	N.A.
California	Travelers[a]	50	Chubb/Pacific Indemnity
	Argonaut[a]	15	Signal/Imperial
	CNA[a]	10	
Colorado	Hartford[a]	55	Empire Casualty
Connecticut	Aetna C&S[a]	80	N.A.
Delaware	Aetna C&S[a]	50	N.A.
District of Columbia	Hartford[a]	80	N.A.
Florida	Argonaut[a]	80	N.A.
Georgia	St. Paul Fire & Marine[a]	90	N.A.
Hawaii	Argonaut[a]	80	N.A.
Idaho	Argonaut[a]	60	N.A.
Illinois	Medical Protective	50	Hartford[a]
Indiana	Medical Protective	60	St. Paul
Iowa	Medical Protective	60	St. Paul
Kansas	Medical Protective	60	St. Paul
Kentucky	Medical Protective	60	Aetna C&S
Louisiana	Hartford[a]	35	St. Paul
Maine	Hartford	45	N.A.
Maryland	St. Paul Fire & Marine[a]	90	N.A.
Massachusetts	St. Paul Fire & Marine	60	Medical Protective Argonaut[a]
Michigan	Medical Protective	68	Shelby Mutual
Minnesota	St. Paul Fire & Marine[a]	90	N.A.
Mississippi	St. Paul Fire & Marine[a]	60	N.A.
Missouri	Medical Protective	60	St. Paul
Montana	Aetna C&S[a]	N.A.	N.A.
Nebraska	Medical Protective	50	St. Paul[a]
Nevada	Argonaut[a]	50	Signal/Imperial
New Hampshire	Hartford	75	N.A.
New Jersey	Chubb/Pacific Indemnity[a]	75	N.A.
New Mexico	Travelers[a]	90	N.A.
New York	Argonaut[a]	80	N.A.
North Carolina	St. Paul Fire & Marine[a]	75	N.A.
North Dakota	St. Paul Fire & Marine	50	N.A.
Ohio	Medical Protective	60	Shelby Mutual
Oklahoma	INA[a]	80	N.A.
Oregon	CNA[a]	60	St. Paul
Pennsylvania	Medical Protective	50	Argonaut[a]
Rhode Island	St. Paul Fire & Marine[a]	N.A.	N.A.
South Carolina	St. Paul Fire & Marine[a]	60	Aetna C&S
South Dakota	St. Paul Fire & Marine	65	N.A.
Tennessee	Shelby Mutual[a]	80	St. Paul
Texas	Medical Protective	60	St. Paul
Utah	Aetna C&S[a]	N.A.	St. Paul
Vermont	Aetna C&S[a]	N.A.	N.A.
Virginia	St. Paul Fire & Marine	60	N.A.
Washington	Aetna C&S[a]	N.A.	N.A.
West Virginia	Aetna C&S[a]	N.A.	N.A.
Wisconsin	Medical Protective	45	St. Paul(35%)
Wyoming	Aetna C&S[a]	N.A.	N.A.

SOURCE: B. Steves and A. McWhorter, Jr., "Notes on the Malpractice Insurance Market," *CPCU Annals*, Vol. 28, No. 4, December 1975, p. 225.

[a] Program with state or local medical societies in existence. In addition to these state programs, Chubb/Pacific Indemnity writes osteopaths countrywide for the American Osteopathic Association (AOA), a program for obstetricians and gynecologists, and a program for the American Psychiatric Association (APA). At the end of 1974, Liberty Mutual still wrote the American College of Physicians (ACP).

Table A.2

ISO EXCESS LIMIT TABLES FOR PHYSICIANS AND SURGEONS, 1955 to 1976

Liability Limits (000)	1955-1968	1969	1970 to 1974		1975		1976	
			Physicians	Surgeons	Physicians	Surgeons	Physicians	Surgeons
5/15	1.00	1.00	1.00	1.00	--	--	--	--
10/30	1.35	1.53	1.66	1.93	--	--	--	--
15/45	1.55	1.83	2.04	2.45	--	--	--	--
20/60	1.64	1.96	2.20	2.68	--	--	--	--
25/75	1.71	2.07	2.34	2.87	1.00	1.00	1.00	1.00
30/90	1.77	2.16	2.45	3.03	--	--	--	--
50/150	1.89	2.34	2.68	3.35	1.26	1.27	1.39	1.40
100/300	2.06	2.59	2.99	3.78	1.49	1.52	1.74	1.77
200/600	2.19	2.79	3.24	4.13	1.66	1.71	1.99	2.05
250/750	2.23	2.85	3.31	4.24	1.72	1.78	2.09	2.16
500/1500	2.36	3.04	3.55	4.57	1.91	1.95	2.37	2.41
1000/3000	2.57	3.26	3.95	5.13	2.21	2.28	2.82	2.90
1500/4500	2.72	3.58	4.23	5.52	2.42	2.49	3.14	3.21
2000/6000	2.83	3.75	4.44	5.81	2.57	2.65	3.37	3.45
2500/7500	2.92	3.88	4.60	6.04	2.69	2.78	3.35	3.64
3000/9000	2.99	3.99	4.74	6.23	2.80	2.89	3.71	3.81

SOURCE: B. Steves and A. McWhorter, Jr., "Notes on the Malpractice Insurance Market," *CPCU Annals*, Vol. 28, No. 4, December 1975, p. 227.

Table A.3

DETERMINANTS OF MALPRACTICE PREMIUMS
DEPENDENT VARIABLE: PREMIUM ($\log_e$)

(t-statistics in parentheses)

Variable	1976 (n=2368)	1975B (n=2251)		1975A[a] (n=753)	1974[a] (n=735)
Specialty					
GP etc.[b,c]	.128	.162	IM[d]	-.244	-.343
	(2.48)	(2.89)		(-3.40)	(-4.48)
General surgery	1.431	1.507	PED	-.420	-.503
	(20.58)	(19.98)		(-5.09)	(-5.69)
OBG	1.565	1.659	OBG/GS	1.349	1.194
	(22.97)	(22.25)		(15.37)	(12.80)
Neurosurgery	1.768	1.735			
	(18.59)	(16.73)			
Ophthalmology	.886	.943			
	(10.29)	(10.05)			
Orthopedic surgery	1.738	1.825			
	(18.72)	(18.16)			
ENT	1.337	1.390			
	(15.47)	(14.76)			
Psychiatry	-.603	-.578			
	(-10.06)	(-8.91)			
Urology	1.339	1.44			
	(14.56)	(14.31)			
Practice Variables and Liability Limits					
MDs	.031	.026		.002	-.003
	(2.38)	(1.83)		(.12)	(-.15)
Other employees	.030	.028		.040	.040
	(5.17)	(4.38)		(4.14)	(2.53)
GP ops.[b]	.264	.241			
	(8.03)	(6.80)			
Total limit	.098	.099		.151	.095
	(2.92)	(2.74)		(4.34)	(.05)
Claim limit	.118	.118		-.084	.022
	(3.95)	(3.72)		(-2.17)	(2.19)

Table A.3--continued

Variable	1976 (n=2368)	1975B (n=2251)	1975A[a] (n=753)	1974[a] (n=735)
Practice Variables and Liability Limits				
Umbrella	.161	.172	.143	.160
	(6.96)	*(6.77)*	*(2.05)*	*(2.19)*
Basic loss	.086	.049	.237	.100
	(1.17)	*(.61)*	*(.711)*	*(.281)*
Basic growth	.339	.369	.485	.473
	(5.33)	*(5.41)*	*(3.39)*	*(3.06)*
Frequency	.384	.404	.369	.531
	(8.01)	*(7.74)*	*(3.62)*	*(4.88)*
Frequency growth	.025	.020	-.150	-.168
	(2.93)	*(2.14)*	*(-4.00)*	*(-4.21)*
Excess ratio	.412	.492	-.068	-.455
	(2.12)	*(2.32)*	*(-.17)*	*(-1.06)*
Other State Variables				
Pool	.020	.031	.193	.245
	(1.07)	*(1.57)*	*(3.92)*	*(4.55)*
Discovery	.194	.153	.264	.112
	(4.12)	*(2.98)*	*(2.26)*	*(.90)*
Program	.094	.094	.110	.166
	(2.27)	*(2.09)*	*(1.53)*	*(2.16)*
Program × surgeon	-.173	-.122	-.294	-.278
	(-2.95)	*(-1.93)*	*(-2.77)*	*(-2.47)*
C	7.141	6.984	5.363	5.744
	(30.77)	*(26.92)*	*(4.09)*	*(4.16)*
R^2	.6080	.6062	.5915	.5556

[a] 21 states only.

[b] GP etc. includes general practice, allergy, cardiovascular disease, gastroenterology, internal medicine.

[c] Omitted category is pediatrics.

[d] Omitted category is general practice.

Table A.4

CONTRIBUTION OF STATE LOSS VARIABLES, NUMBER OF PHYSICIANS AND DISCOVERY RULE

Equation Number	Variables	1976			1975B		
		DF	SS	R^2	DF	SS	R^2
1	L, R	7	143.1	.050	7	166.8	.056
2	L	5	136.2	.048	5	161.8	.054
3	(1-2) R	2	6.9	.002	2	5.0	.002
4	R	2	58.5	.021	2	63.9	.021
5	(1-4) L		84.6	.029	5	102.9	.035
Residual		2361	2964.7		2244	2822.7	
Total		2368	2837.8		2251	2989.5	

L = Loss characteristics (basic loss, frequency, excess ratio, basic growth, frequency growth).

R = Risk (discovery, pool).

Table A.5

CONTRIBUTION OF MEASURED STATE (Z), PROGRAM AND INTERACTION (P, P×S), NET OF MEASURED PHYSICIAN (X)

Equation Number	Variables	1976[a]			1975B[a]			1975A[b]			1974[b]		
		DF	SS	R^2	DF	SS	R^2	DF	SS	R^2	DF	SS	R^2
1	X, Z, P, P×S	18	1669.8	.588	18	1760.0	.589	17	477.0	.592	17	443.1	.556
2	X	9	1509.4	.532	9	1599.3	.535	8	375.6	.466	8	324.1	.406
3	(1-2) Z, P, P×S	9	160.4	.057	9	160.7	.054	9	101.4	.126	9	119.0	.150
4	X, Z	16	1666.0	.587	16	1757.6	.588	15	473.6	.587	15	439.5	.551
5	(1-4) P, P×S	2	3.8	.001	2	2.4	.001	2	3.4	.005	2	3.6	.005
6	X, Z, P	17	1666.2	.587	17	1758.3	.588	16	473.6	.587	16	440.1	.552
8	(1-6) P×S	1	3.6	.001	1	1.7	.001	1	3.4	.005	1	3.0	.004
Residual		2350	1168.0		2233	1229.5		736	329.4		718	354.4	
Total		2368	2837.8		2251	2989.5		753	806.4		735	797.5	

[a] 50 states.

[b] 21 states.

X = Specialty, practice size, coverage limits.
Z = Loss characteristics, pool, discovery.
P = Program.
P×S = Program × surgeon.

Table A.6

CONTRIBUTION OF MEASURED PHYSICIAN (X) AND INTERACTION (P×S), NET OF STATE (D)

Equation Number	Variables	1974[b]			1975A[b]			1975B[a]			1976[a]		
		DF	SS	R^2	DF	SS	R^2	DF	SS	R^2	DF	SS	R^2
1	X, D, P×S	29	496.7	.623	29	525.1	.651	59	1907.2	.6380	59	1781.9	.628
2	D	20	225.5	.283	20	204.2	.253	49	323.3	.108	49	263.4	.093
3	(1-2) X, P×S	9	271.2	.340	9	320.9	.398	10	1583.9	.530	10	1518.5	.535
4	X, D	28	494.4	.620	28	522.4	.648	58	1906.0	.6376	58	1779.5	.627
5	(1-4) P×S	1	2.3	.003	1	2.7	.003	1	1.2	.0004	1	2.4	.001
6	(3-5) X	8	268.9	.337	8	318.2	.395	9	1582.7	.530	9	1515.1	.534
Residual		706	300.8		724	281.3		2192	1082.3		2309	1055.9	
Total		735	797.5		753	806.4		2251	2989.5		2368	2837.8	

[a] 50 states.

[b] 21 states.

X = Specialty, practice size, coverage limits.
Z = Loss characteristics, pool, discovery.
P = Program.
P×S = Program × surgeon.
D = Vector of state dummies.

Table A.7

DECOMPOSITION OF STATE (D) INTO MEASURED (Z) AND UNMEASURED (D-Z), NET OF MEASURED PHYSICIAN (X)

Equation Number	Variables	1974[b]			1975A[b]			1975B[a]			1976[a]		
		DF	SS	R^2	DF	SS	R^2	DF	SS	R^2	DF	SS	R^2
1	X, D, D×S	47	509.9	.639	47	535.6	.664	107	1958.7	.655	107	1823.7	.643
2	X	8	321.1	.406	8	375.6	.466	9	1599.3	.535	9	1509.4	.532
3	(1-2) D, D×S	39	188.8	.233	39	160.0	.198	98	359.4	.120	98	314.3	.111
4	X, D	28	494.4	.620	28	522.4	.648	58	1906.0	.638	58	1779.5	.627
5	(1-4) D×S	19	15.5	.019	19	13.2	.016	49	52.7	.017	49	44.2	.016
6	X, Z, P, P×S	17	443.1	.556	17	477.0	.592	18	1760.0	.589	18	1669.8	.588
7	(1-6) [D, D×S] - [Z, P, P×S]	30	66.8	.083	30	58.6	.072	89	198.7	.066	89	153.9	.055
8	X, D, P×S	29	496.7	.623	29	525.1	.651	59	1907.2	.638	59	1781.9	.628
9	(1-8)[D, D×S] - [D, P×S]	18	13.2	.016	18	10.5	.013	48	51.5	.017	48	41.8	.015
Residual		688	287.6		706	270.8		2144	1030.8		2261	1014.1	
Total		735	797.5		753	806.4		2251	2989.5		2368	2837.8	

[a] 50 states.

[b] 21 states.

X = Specialty, practice size, coverage limits.
Z = Loss characteristics, pool, discovery.
P = Program.
P×S = Program × surgeon.
D = Vector of state dummies.

Table A.8

ANALYSIS OF 1975 ISO RECOMMENDED RATES

Equation Number	Variables	DF	SS	R^2
1	L, P, R	8	5.67	.387
2	L, R	7	5.04	.344
3	(1-2) P	1	.63	.043
4	L, P	6	5.02	.343
5	(1-4) R	2	.65	.044
Residual		37	8.97	
Total		45	14.63	

L = Loss characteristics.
P = Program.
R = Risk (discovery, pool).

Table A.9

REGRESSIONS OF 1975 ISO RECOMMENDED RATE INCREASE

(t-statistics in parentheses)

ISO rate	-.218**	-.200*	-.209**
	(-2.13)	(-1.93)	(-2.02)
ISO pending	.713**	.630**	.717**
	(3.42)	(2.96)	(3.39)
ISO loss ratio	14.330	16.763*	
	(1.46)	(1.78)	
Frequency growth	.154		.217
	(.925)		(1.32)
Frequency	.030*		.027*
	(1.85)		(1.65)
Excess ratio	.031		.031
	(1.40)		(1.66)
C	-7.44	-9.13	2.088**
	(-1.14)	(-1.46)	(3.13)
R^2	.468	.388	.439
n = 45			

[*] Significant at the 10 percent level.

[**] Significant at the 5 percent level.

BIBLIOGRAPHY

American Medical Association, *Physician Distribution and Medical Licensure in the U.S., 1974,* Chicago, 1975.

Brown, John P., "Toward an Economic Theory of Liability," *Journal of Legal Studies,* Vol. 2, 1973.

Greenspan, Nancy T., "A Descriptive Analysis of Medical Malpractice Insurance Premiums, 1974-77," *Health Care Financing Review,* Fall 1979.

Insurance Services Office (ISO), *Compilation of General Liability Experience: Professional Liability,* New York, 1975.

Kendall, Mark and John Haldi, "The Medical Malpractice Insurance Market," in Appendix to *Report of the Secretary's Commission on Medical Malpractice,* DHEW Publication No. (OS) 73-89, January 1973.

Munch, Patricia, "Causes of the Medical Malpractice Insurance Crisis: Risks and Regulation," in Simon Rottenberg (ed.), *The Economics of Medical Malpractice,* American Enterprise Institute, Washington, D.C., 1978.

Munch, Patricia, *The Disposition of Medical Malpractice Claims,* The Rand Corporation, R-2622-HCFA, October 1980.

Munch, Patricia and Dennis E. Smallwood, "Professional Liability," Staff background paper prepared for the California Citizens' Commission on Tort Reform, February 1979.

National Association of Insurance Commissioners (NAIC), *Malpractice Claims,* Vol. 1, No. 4, May 1977.

Posner, Richard, "A Theory of Negligence," *Journal of Legal Studies,* Vol. 1, 1972.

Shavell, Steven, "Accidents, Liability and Insurance," Discussion Paper Number 685, Harvard Institute of Economic Research, Harvard University, Cambridge, February 1979.

Steves, Myron and A. McWhorter, Jr., "Notes on the Malpractice Insurance Market," *CPCU Annals,* Vol. 28, No. 4, December 1975.

Westat, Inc., *Study of Medical Malpractice Claims Closed in 1970,* Rockville, Maryland, 1973.

THE DISPOSITION OF MEDICAL MALPRACTICE CLAIMS

PREPARED FOR THE HEALTH CARE FINANCING ADMINISTRATION,
U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

PATRICIA MUNCH DANZON

PREFACE

The research reported here was funded under a contract with the U.S. Department of Health, Education, and Welfare, Contract No. 600-76-0150, "An Economic Analysis of Medical Malpractice." Other work performed under this contract includes *Contingent Fees for Personal Injury Litigation* (R-2458-HCFA), *Why Are Malpractice Premiums So High--or So Low?* (R-2623-HCFA), and *The Determinants of the Frequency of Malpractice Claims* (R-2661-HCFA/ICJ).

This report develops a model of the disposition of malpractice claims: the size of awards (at verdict and in out-of-court settlements), the probability that some positive payment is made to the plaintiff, and the propensity to settle or litigate to verdict. The model assumes that the courts apply the legal standards of liability and damages, subject to information provided by the litigants. The litigants decide how much to spend and when to settle, based on their expectations of the payoff and cost of litigating further. If these assumptions are correct, the disposition of claims is predicted to conform to some extent but also to depart systematically from the legal standard of payment equal to damages if and only if negligence occurred. The model is used to structure the analysis of data from three surveys of malpractice claims, closed in 1970, 1974, and 1976. Estimates are presented of the effects on the disposition of claims of characteristics of the injury, the plaintiff, the defendants, and the state in which the injury occurred.

The estimates reported here are reduced form estimates conditional on the stage of disposition. Estimation of the structural model is the subject of ongoing research with Lee Lillard.

SUMMARY

This report compares the actual outcome of medical malpractice claims with the legal standard of payment equal to damages if and only if negligence occurs. A theoretical model of the disposition of claims extends earlier models of the settlement process and assumes that the courts apply the legal standards of liability and damages, but depend on information provided by the litigants. The litigants are assumed to maximize wealth, given their expectations of the behavior of the courts and the costs of litigating. Because litigation is costly, both parties can gain if they can agree to settle. The plaintiff forms a minimum "ask" equal to his expected net gain from further litigation. If expected payoff exceeds the costs of litigating, the case may be dropped. Similarly, the defendant forms a maximum offer equal to his total expected payout. If the offer exceeds the ask, for some amount within this bargaining range there could be a settlement. If the ask is positive and exceeds the offer, litigation continues.

If the assumptions of the model are correct, the disposition of claims is expected to depart systematically from the legal standard of awards equal to damages if and only if negligence occurred. Specific predictions include the following:

- The probability that the plaintiff will win at verdict and the size of award will be positively (negatively) related to the cost of litigation for the defense (plaintiff).
- Cases settled out of court will be influenced by legal standards as reflected in the litigants' expectations of the verdict if the case went to court. However, settlements will discount the expected verdict by the probability that the plaintiff will win. Settlements will also be positively related to the defendant's litigation costs and errors in predicting the outcome, and negatively related to the plaintiff's litigation costs and errors in predicting the outcome.

- The probability of payment to the plaintiff in settlement depends on the probability that the plaintiff will win at verdict and, hence, on the legal standards and evidence of negligence. However, it will also be positively related to the expected award and to the litigation costs of the defense, and negatively related to litigation costs of the plaintiff.
- A sufficient condition for litigation is that the plaintiff's expectation of the outcome from continuing exceed the defendant's by more than the sum of their litigation costs. This implies that the greater the expected award and the uncertainty as to the outcome, and the lower the cost of litigating for either side, the higher the probability of litigation.
- The sample of cases closing at each stage of disposition will not be a random sample of the claims population. In particular, cases tried to verdict will tend to be the least predictable, hence they are probably atypical.

Estimation of the determinants of probability of winning and size of award, by stage of disposition, yields conditional estimates for the stage of disposition but biased estimates of population parameters. Conditional estimates are reported here. Unconditional estimates are the subject of ongoing research with Lee Lillard. Data from three surveys of malpractice claims closed in 1970, 1974, and 1976 were used. Because of the limited quality of the data, conclusions are tentative. However, the results are generally consistent with the legal standard and the economic model.

Awards are strongly related to the severity of the injury and, for permanent injuries, to the life expectancy of the plaintiff. Consistent with the law of compensable damages, cases involving death receive substantially less than cases of permanent disability. Awards are positively related to earnings for persons employed at the time of the injury. On average, employment status and sex do not have a significant effect.

The adequacy of compensation was not measured, but a more limited hypothesis was measured--that awards increase at least in proportion to pecuniary loss. The hypothesis of proportionality implies that the result reported by others--that awards tend to be greater than losses for cases involving fairly low losses and less for cases involving fairly large losses--could be due entirely to error in reported losses.

Awards depart from the legal standard in ways predicted by the economic model of disposition, that settlements discount expected court awards by the probability that the plaintiff will win. Proxies for that probability (insurer's evaluation of liability, allegation of *res ipsa*, representation by an attorney) are all associated with higher settlements.[1] As predicted, settlements are also positively related to the litigation costs of the defense; the elasticity of award with respect to number of hospital defendants is .2; for physician defendants it is .4.[2] Settlements are negatively related to the litigation costs of the plaintiff, as represented by the lag in reporting the claim, for cases settled at suit.[3] Awards are positively related to the limits of the defendants' liability insurance coverage, after number of defendants is controlled for.

Measurement of the determinants of the probability that the plaintiff will win was generally less successful than measurement of size of awards. However, what little information there is on the evidence of negligence is consistent with the legal standard. The single most powerful predictor of payment being made to the plaintiff, at verdict and in settlement, is the insurance company's evaluation that the claim had legal merit. Cases involving obvious error or alleging *res ipsa* have a higher probability of winning, as do cases where the injury was induced by treatment. Cases involving misdiagnosis or lack of informed consent, incomplete cure or unauthorized treatment are less likely to receive payment. Representation by an attorney increases the probability of payment.[1]

---

[1] The positive association between attorney representation and settlements probably reflects imperfectly measured loss and positive marginal product of the attorney as well as a higher probability of winning.

[2] The higher value for physicians is attributed to higher time and embarrassment costs of going to court.

[3] This coefficient may overestimate the effect of plaintiff costs because of correlation with imperfectly measured loss and evidence of negligence.

The model predicts that the probability of payment to the plaintiff depends not only on the evidence of negligence but also on the size of the expected award and the litigation costs of the defense and (negatively) on the litigation costs of the plaintiff. The probability of out-of-court payment to the plaintiff is positively related to the severity and duration of the injury and to reported loss (proxy for expected award), and negatively related to the lag in reporting the claim (proxy for plaintiff litigation costs). However, the effect of number of defendants (proxy for defense litigation costs) is significantly positive only in 1974. Because roughly 50 percent of all claims receive no payment, the allegation that insurance companies can be forced to pay on all claims, no matter how "frivolous," is clearly false.

The propensity to litigate is positively related to the expected award, negatively related to the costs of litigation, and positively related to uncertainty as to the outcome. The data available explain less of the decision to go to verdict than the decision to file suit. The size of the compensable damages has the expected positive effect. Strong evidence of negligence tends to encourage settlement. Cases involving multiple defendants are more likely to go to suit. The tendency for multiple defendants to cause costs to be higher for the defense and thereby encourage settlement is more than offset by the increase in uncertainty and possibly lower costs for the plaintiff.

Measuring the effect of state characteristics was generally unsuccessful. Individual coefficients were sensitive to the specification and unstable across years, possibly because of correlation with other measured and unmeasured variables, imprecise measurement, simultaneous determination, or simply no effect. The 1975-76 changes in malpractice law were enacted after many of the claims used here had been filed and it was not possible to measure full, long-run effects. Even as a group, state characteristics added little explanatory power. However, the strong positive effects on the propensity to litigate of the number of claims in the state, the excess limits ratio, and the average plaintiff win rate suggest that distinct differences across states are not captured by the variables used in this analysis.

---

[1] One measure of the effect of having an attorney is downward biased because it is conditional on closing at pre-suit. The higher probability of winning at suit than pre-suit is in part due to attorney representation, which is virtually 100 percent at suit.

ACKNOWLEDGMENTS

I am indebted to Christine Peterson and Richard Buddin for very able research assistance. Rand colleagues John Cogan and Emmett Keeler provided many useful comments on an earlier draft. The American Insurance Association kindly provided the data.

This study has also benefited from the related, ongoing research with Lee Lillard.

CONTENTS

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS xi

Section

I. INTRODUCTION 1

II. MODEL OF THE DISPOSITION OF MALPRACTICE CLAIMS 4
  Overview 4
  Formal Statement of the Model 5

III. IMPLICATIONS OF THE MODEL 11
  Awards 11
  Payment to the Plaintiff 11
  Expenditure on Litigation 12
  Selection by Stage of Disposition 13

IV. THE DATA 15

V. SIZE OF AWARDS 23
  The Legal Standard 23
  Implications of the Model 24
  Findings by Others 24
  Empirical Estimates 25

VI. THE PROBABILITY THAT THE PLAINTIFF WINS 44
  Implications of the Model 44
  Empirical Estimates 45

VII. THE DECISION TO LITIGATE 53
  Implications of the Model 53
  Empirical Estimates 53

VIII. CONCLUSION 60

APPENDIX 65

BIBLIOGRAPHY 71

I. INTRODUCTION

Since 1975, the majority of states have enacted new laws on medical malpractice liability. The specifics of the legislation vary from state to state, but the common purpose was to hold down the cost of claims and, hence, insurance.[1] The legislation modified the scope of liability, the size and form of payment, the judicial process for handling disputed claims, and malpractice insurance arrangements. These changes were enacted in an atmosphere of crisis, with little analysis of the merits and defects of the existing system or the probable effects of the changes. The research reported here provides a theoretical and empirical analysis of the disposition of malpractice claims as one input into an evaluation of the current system and proposed reforms.

According to the law defining liability for medical malpractice, a physician is negligent if and only if he departs from the standard of care. An award equal to the damages suffered is then made to the injured party. These principles of negligence law define rules that, under certain conditions, create incentives for the efficient level of investment in injury prevention.[2] Sufficient conditions for the negligence system to induce efficient investment in injury prevention (optimal deterrence) are: (1) The standard of negligence applied by the courts should be the economically efficient standard; (2) payment equal to the damages caused should be made by the negligent party in all cases of departure from the standard; (3) the system should be costless to operate, which requires that the courts have perfect information about the behavior of the potential tortfeasor and the potential tortfeasor have perfect information about the standard applied by the courts.

---

[1] Malpractice insurance premiums increased over 300 percent in some states in 1974-75.

[2] For example, Posner (1972), Brown (1973). Efficiency requires preventing an injury if the cost of prevention is less than the expected damages, i.e., the probability that the injury will occur multiplied by the loss suffered if it does. This is equivalent to the legal definition of negligence laid down by Judge Learned Hand in U.S. vs. Carroll Towing Co., 159 Federal Reporter 2nd 169 (1947).

In practice, the medical malpractice system is alleged to violate the legal standard and therefore the conditions for economic efficiency. First, it is widely believed that payment is made in many cases that do not involve negligence, either because of jury sympathy for a severely injured plaintiff or because insurance companies can be forced to pay on claims with little legal merit to avoid costly litigation. Second, awards are variously believed to be too high, too low, or random relative to damages suffered. In 1976, 5 percent of claims accounted for 73 percent of total indemnity paid to all plaintiffs (NAIC, 1977). However, an earlier study (AIA, 1976) of the relation between awards and "economic loss" (which includes only wage loss and medical and other tangible expenses) found that awards on average are less than economic loss and the larger the economic loss the greater the shortfall. If so, awards are *a fortiori* less than full damages, including intangible components of pain and suffering. However, several states have enacted ceilings on awards, indicating a belief that large awards are excessive.

Third, the system is far from costless to operate; rough estimates based on the limited data available show that only 30 to 40 cents of every malpractice premium dollar reaches the plaintiff as compensation. An equivalent amount is spent on litigation.[1]

This report develops a theoretical model of claim disposition, based on the assumption that the courts do not have perfect information and the parties to the claim attempt to maximize wealth, given their expectations of the behavior of the courts and the costs of litigating. The purpose of the model is to show how these assumptions lead to outcomes that depart from the legal standard of payment equal to damages if and only if negligence occurred and to show how the characteristics of the plaintiff, the defense, and the case affect all aspects of the outcome: the probability that the plaintiff wins, the size of payment, and the stage of disposition (pre-suit, at suit, or by trial to verdict). Evidence from three surveys of malpractice claims closed in 1970, 1974, and 1976 is used to address the following questions:

---

[1] Munch (1977a). This understates the total cost of the system because it excludes the cost of operating the courts. It is argued below that expenditure on litigation is evidence of uncertainty. Uncertainty is costly not only because it induces expenditure on litigation but also because it undermines the deterrent function of the tort system. If the legal standard of negligence is unclear, it cannot provide the physician with a clear signal as to the appropriate standard of care. Uncertainty also creates incentives for the physician to protect himself by practicing defensive medicine or by purchasing liability insurance rather than by attempting to avoid negligent practice. Liability insurance in turn dulls the incentive to practice with care (unless premiums are perfectly experience-rated), thereby further undermining the deterrent function of the tort system.

- To what extent do outcomes conform to the legal standard?
- Are departures from the legal standard consistent with the model of disposition?
- What is the *net* effect of specific case characteristics? Do the simple correlations reported by others persist in multivariate analysis?
- How do specific laws and other state characteristics affect the outcome?
- What are the implications for tort reform?

II. MODEL OF THE DISPOSITION OF MALPRACTICE CLAIMS

OVERVIEW[1]

The model assumes that the courts apply the rules for determining liability and damages but lack perfect information about the facts of the case. Decisions are therefore based on the evidence provided by the litigants.

The plaintiff and defendant are assumed to maximize their expected wealth: The plaintiff maximizes his expected recovery, net of legal costs; the defendant minimizes his expected payout, gross of legal costs.[2] Maximization implies two related choices. First, one must determine the optimum (potential maximum) expenditure on litigation, which depends on the expected payoff (in terms of influencing the outcome in court) relative to the cost of litigation inputs. Because litigation is costly, both parties can potentially gain if they can agree to settle. Therefore, the second decision is whether to settle, litigate further, or (for the plaintiff) drop the case. Maximization implies that the plaintiff determines a minimum ask (the least amount he will settle for) equal to his expected payoff from litigating further, net of his incremental litigation costs. Similarly, the defendant's maximum offer is his expected payout from litigating further, gross of his incremental litigation costs. (Hereafter, "ask" and "offer" refer to these extremes of the bargaining range.) The ask and offer may be revised with changes in the optimum incremental expenditure and expectations of outcomes as negotiations proceed.

---

[1] A similar model of the settlement process is developed in Gould (1973), Landes (1971), Posner (1973), Munch (1976).

[2] The model ignores possible conflict of interest among multiple defendants and between principals and agents--the defendant and his insurance company and the plaintiff and his attorney. Conflict of interest among multiple defendants is discussed in Munch (1977b). The contingent fee contract typically used by plaintiff attorneys in personal injury cases is analyzed in Munch (1980) where it is shown that under plausible assumptions, the contingent fee contract induces the attorney to act in his client's best interest, contrary to conclusions reached by others (e.g., Schwartz and Mitchell, 1970). Thus the neglect here of the agency problem may be reasonable.

A necessary condition for the case to be dropped without payment is that the ask becomes negative. If the ask is positive but less than the offer, the case may be settled for some positive amount within the range between them.[1] If the ask exceeds the offer, litigation continues. Actual expenditure on litigation may therefore be less than the potential maximum. A necessary condition for continuing to litigate is that the plaintiff's expected payoff exceeds the defendant's by more than the sum of their incremental expenditure on litigation.

This model of the disposition process and the percent of claims closed at each stage of disposition are illustrated in Fig. 1. A formal statement of the model is provided below. The reader who is not interested in it can skip to Sec. III, which summarizes the implications for the disposition of cases (the size of verdicts and settlements, the probability the plaintiff obtains some positive recovery, expenditure on litigation and the stage of disposition) and interpretation of data from claims files.

FORMAL STATEMENT OF THE MODEL

The following notation is used:

V = award at verdict.

q = probability of a verdict for the plaintiff.

Z_1 = vector of characteristics affecting V (severity of injury, plaintiff's life expectancy, income, etc.).

Z_2 = vector of characteristics affecting q (evidence of negligence, rules of liability, etc.),

L = litigation inputs (time of attorneys, witnesses, litigants, etc.).

w = average cost per unit of litigation inputs (weighted average of wage rate of attorneys, value of time of plaintiff or defendant, etc.).

---

[1] The ask being less than the offer is a necessary but not a sufficient condition for settlement, because the actual ask and offer may differ from these hypothetical bounds in an attempt to influence the actual settlement amount. The strategy of the bargaining process is not modeled explicitly. Similarly, a negative ask is a necessary but not a sufficient condition for the case to be dropped.

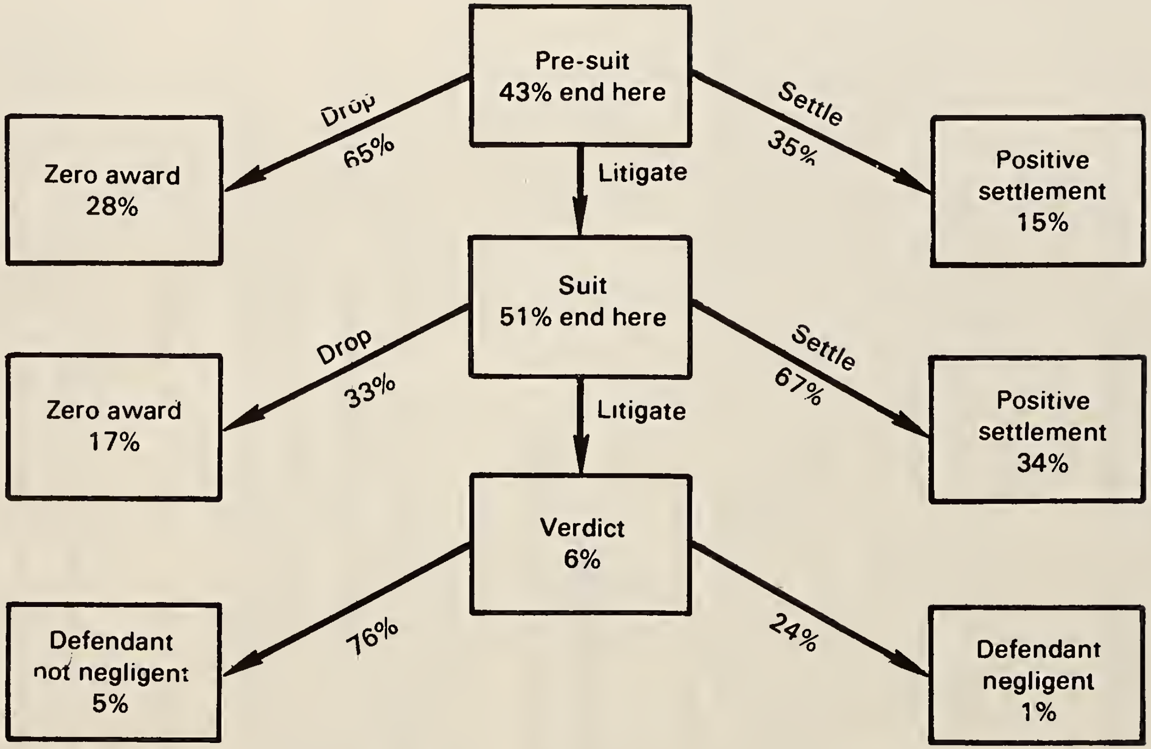

Fig. 1 — Disposition of claims

S = settlement.

B = litigants' expectation of V, conditional on a positive award.

r = litigants' expectation of q.

A = minimum ask or maximum offer in pre-verdict bargaining.

Superscript p denotes plaintiff.

Superscript d denotes defendant.

A circumflex (^) denotes optimized value.

NOTE: In a more general model, the expectations that the plaintiff will win (r) and size of payment (B) could be interpreted as expectations over the distributions of possible outcomes--settlement or trial to verdict.

Courts

The probability of a verdict for the plaintiff depends on the evidence of negligence (Z_2) and the effort of the plaintiff and defendant to influence the courts (L):

$$q = q(Z_2, L^p, L^d). \tag{1}$$

The size of the award (V), if positive, depends on compensable damages and other factors (Z_1) and the effort of the litigants:

$$V = V(Z_1, L^p, L^d). \tag{2}$$

Choice of Inputs

The plaintiff and defendant determine L given the cost per unit (w) and the expected payoff, rB(L).[1]

Plaintiff. The plaintiff chooses L^p to maximize his expected net recovery:

$$\max_{L^p}: \quad r^p B^p - wL^p,$$

---

[1] The payoff function rB(L) is assumed to have the properties necessary for a unique interior maximum.

yielding

$$\frac{\partial (rB)^p}{\partial L^p} = w. \tag{3}$$

Defense. The defendant chooses L^d to minimize his expected payout, gross of litigation expense:

$$\min_{L^d}: \quad r^d B^d + wL^d,$$

yielding

$$-\frac{\partial (rB)^d}{\partial L} = w. \tag{4}$$

The optimized values of inputs and implied optimized expected outcomes are denoted $\hat{L}$, $\hat{r}$ and $\hat{B}$. These optimized values may be revised as more information accrues about the facts of the case and behavior of both sides. Time subscripts are deleted for expository convenience.

The Bargaining Range

At any point before the verdict, the plaintiff's minimum ask (A^p) is his expected (optimized) payoff from litigating further, net of his incremental litigation costs:

$$A^p = (\widehat{rB})^p - w\hat{L}^p. \tag{5}$$

Similarly, the defendant's maximum offer (A^d) is his expected (optimized) payout plus his incremental litigation costs:

$$A^d = (\widehat{rB})^d + w\hat{L}^d. \tag{6}$$

Note that A^p and A^d represent the extreme positions that each party would be willing to settle for and may be concealed from the other side for strategic reasons.[1]

The Outcome

Case Dropped (Zero Payment). If the ask becomes negative (i.e., the cost of litigating further exceeds the expected payoff), the case may be dropped with zero payment:

$$A^p < 0 \rightarrow (\widehat{rB})^p < w\hat{L}^p \rightarrow \text{Drop Possible.} \tag{7}$$

Settlement. If the minimum ask is positive but less than the maximum offer, settlement can occur:

$$0 < A^p \leq A^d \rightarrow \text{Settlement Possible.} \tag{8}$$

The amount of settlement is some weighted average of the ask and offer:

$$S = gA^d + (1 - g)A^p. \tag{9}$$

Litigation. If the ask is positive and exceeds the offer, litigation continues:

$$0 < A^p \text{ and } A^d < A^p \rightarrow \text{Continue Litigation.}$$

Estimation

The full structural model cannot be estimated because the plaintiff's expenditure, the ask, the offer, the probability of winning and the expected award at verdict for cases settled before verdict are not observed. The model to be estimated consists of reduced form equations for awards at verdict and settlement (at suit and at pre-suit);

---

[1] Plaintiff risk aversion (preference) lowers (raises) A^p; defendant risk aversion (preference) raises (lowers) A^d.

the probability that the plaintiff wins at verdict, suit, and pre-suit; and the decision to litigate from pre-suit to suit and from suit to verdict.

III. IMPLICATIONS OF THE MODEL

This model implies that the disposition of malpractice claims will depart in predictable ways from the norm of payment equal to damages if and only if negligence occurred, with zero expenditure on litigation.

AWARDS

Verdicts are determined not only by "true" compensable damages but also by the input of the litigants to influence the courts' perception of those damages. This expenditure on litigation is positively related to the expected payoff and negatively related to the cost per unit. Settlements are influenced by the behavior of the courts, as reflected in the litigants' expectations of the potential verdict. However, settlements discount the expected verdict by the litigants' perception of the probability of a verdict for the plaintiff. Settlements are also positively (negatively) related to the defendant's (plaintiff's) litigation costs and errors in predicting the outcome in court. This follows from the assumption that the settlement is some average of the ask and offer.[1]

PAYMENT TO THE PLAINTIFF

The probability of a verdict for the plaintiff depends not only on the facts of the case but also on the quality of the evidence and on litigation inputs. Payment to the plaintiff at settlement (alternatively, the probability that the case is not dropped) requires that the plaintiff's expected payoff exceed his litigation costs. This incorporates the plaintiff's expectation of the size of the award, the probability of proving negligence, and his costs of litigating. Even if the ask is negative, the case will not be dropped if the defense makes a positive offer. The greater the defendant's expectations of his payout in court and costs of litigating, the more probable a positive offer. Thus, payment may occur even in cases where, if the facts could be known, there was no negligence. The larger the expected award, the defendant's litigation costs, and both parties' errors in predicting the outcome, the greater the probability of payment to the plaintiff.

---

[1] If the parties have unbiased expectations of the outcome in court and split the difference between the ask and offer, then: $S = qV + .5[(wL)^d - (wL)^p]$. If $(wL)^d = (wL)^p$, then $S = qV$.

EXPENDITURE ON LITIGATION

The greater the potential to influence the outcome in its favor, the greater each party's optimum expenditure on litigation. Expenditure is induced by "suggestibility" of the courts, not randomness. If the courts determined the award by picking a number at random, there would be no payoff to expenditure on litigation. However, if the drawing depends on the evidence presented by the litigants, then expenditure on litigation will be positively related to the change in expected outcome per unit of input.

Although maximum potential expenditure depends on the susceptibility of the courts to influence, actual expenditure can fall short of this potential maximum if each party knows the other's intended expenditure and they agree on the expected outcome. A necessary condition for settlement is that the plaintiff's expected payoff from continuing exceeds the defendant's by more than the sum of their incremental litigation costs.[1] Therefore, although randomness of the courts tends to reduce the maximum *potential* expenditure on litigation, randomness increases the probability of disagreement as to the outcome and hence increases the probability that the case will not be settled and the maximum will be spent.

---

[1] If the parties on average correctly anticipate the effect of their expenditure on the outcome, then this model implies that actual expenditures are a lower bound on the gap between their expectations of the outcome with zero expenditure. For example, if the actual settlement is \$100 and both sides spent \$30, the plaintiff must have expected to get no more than \$70 and the defense must have expected to have to pay at least \$130, with zero expenditure, for the actual expenditure of \$30 apiece to have been worthwhile. It is a lower bound for two reasons. First, the expected value of inframarginal units of expenditure (in terms of changing rB) exceeds their cost, if there are diminishing returns. (With indivisibilities this is not necessarily true.) Second, the model presented here assumes independence. If the parties expect their expenditure to invite partially offsetting expenditure by the other side, the divergence in expected payout with zero expenditure must be even greater.

The potential for settlement does not necessarily increase over time. Expectations may converge as litigation proceeds and the parties learn more about the facts of the case and each other's probable strategy. However, the optimal incremental expenditure on litigation also decreases as more costs become "sunk." Thus the potential for settlement increases over time only if the convergence in expectations exceeds the decrease in optimum incremental expenditure.[1]

SELECTION BY STAGE OF DISPOSITION

The model implies that the subsample of cases observed at each possible outcome is not a random sample from the population of claims but is selected on the basis of the parties' prediction errors. Specifically, cases where the plaintiff (defendant) tends to underestimate either the potential award or his probability of winning are less (more) likely to be litigated and more (less) likely to be dropped without payment.

Non-random selection of cases implies that seriatim estimation of the outcome equations (size of award, probability of the plaintiff winning, decision to litigate) yields biased estimates of population parameters. Because of the complexity of the selection process, the standard techniques for correcting for selection bias are inappropriate. The estimates presented here therefore simply ignore the selection bias problem. They are to be interpreted as conditional on selection into a particular outcome.[1]

---

[1] The model assumes that each party determines its optimum input unilaterally, ignoring matching expenditure by the other side. This is less restrictive than it might appear. For example, if the plaintiff expects the defendant to react optimally, the defendant's matching expenditure will equate a decrease in expected payout with the cost per unit. At the margin, the net effect on the offer is zero, because the offer is simply the defendant's expected payout gross of litigation expense. Matching expenditure by the defendant therefore reduces the payoff to plaintiff expenditure. But the offsetting effect is less if the case settles, as do 90 percent of cases, than if it goes to court.

The appendix formally derives the effects on the outcome of changes in the case's characteristics. The results are summarized in Table 1.

Table 1

PREDICTED EFFECTS ON OUTCOMES OF CHANGES IN CASE CHARACTERISTICS

(Variables from the claims files are in parentheses)

Characteristics	V	q	$\hat{L}^p$	$\hat{L}^d$	A^p	A^d	S	q^*	F
Damages (severity, loss)	+	(+)	+	+	+	+	+	+	?
Evidence of negligence (liability, *res ipsa*)		+	(+)	(+)	+	+	+	+	?
Plaintiff's litigation costs (income, lag in reporting)	-	-	-		-		-	-	-
Defense's litigation costs (number of defendants)	+	+		-		+	+	+	-

V = award at verdict.

q = probability of a verdict for the plaintiff.

$\hat{L}^p$ = plaintiff's optimum litigation input.

$\hat{L}^d$ = defense's optimum litigation input.

A^p = minimum ask.

A^d = maximum offer.

q^* = probability that plaintiff wins if case closes in settlement (1 - probability case is dropped).

S = award in settlement.

F = propensity to litigate.

---

[1] Lee Lillard and I are currently working on a specification of the model to obtain unbiased, unconditional estimates of population parameters for the structural equations of the model.

IV. THE DATA

The data are from three surveys of malpractice claims closed by insurers in 1970, 1974, and 1976.

The 1970 survey was conducted in 1972 for the Secretary's Commission on Medical Malpractice.[1] Of approximately 68 firms writing medical malpractice insurance at that time, 26 participated in the survey. The participating firms accounted for roughly 90 percent of all malpractice insurance issued. Thus, the sample is reasonably representative of claims against physicians. However, claims against hospitals are underrepresented, because an unknown proportion of hospitals self-insure.

The sampling procedure was designed to yield a sample of approximately one-fifth of the claim files closed by the participating insurers during 1970. All claims resulting in payment of at least $100,000 were included.[2] Claims resulting in payment of less than $100,000 were sampled at the basic rate of one in five. Incidents involving multiple defendants therefore had a higher probability of selection into the sample. In the analysis here, incidents are weighted by the inverse of their probability of selection into the sample in order to yield a sample that is representative of the population.

The 1974 survey conducted by the American Insurance Association was not structured as a random nationwide sample.[3] Eleven companies participated in the survey. All states are represented but not equally.[4]

---

[1] The sample design, construction of weights, survey instrument and basic characteristics of the data are described in Westat (1973).

[2] All claim files from a few companies with a small number of incidents were also automatically included.

[3] The survey instrument and data are described in AIA (1976).

[4] For example, New York is underrepresented because participating carriers did little business there.

The 1976 survey was confined to the nine companies involved in the largest number of claims closed by private companies in 1976.[1] Data were collected on all claims closed during the four-month period from July 1 through October 31, 1976.

This survey collected extensive information on the nature of the injury. The level of detail reported is not amenable to the techniques of analysis used here, nor is it necessarily relevant to the disposition of the claims. For this analysis, I defined broad categories of injuries that are likely to influence the ease of proving negligence. Dummy variables indicate if any of these (not mutually exclusive) injury categories were mentioned at least once in the file(s) relating to the incident.

Insurance companies typically open a separate claim file for each defendant named in relation to an incident. Because this analysis focuses on the outcome of the incident, claims against multiple defendants but related to the same incident are combined.[2] Incidents that were reported by an insured rather than a plaintiff (preclaims) were excluded from the analysis if no payment was made.[3] Because the surveys were of insurance company files, the data are confined to information available to the companies, which excludes many factors that are relevant to the outcome. Response rates tend to be low on items not routinely collected by the companies. Incidents were excluded from the analysis if data on key variables were missing or inconsistent.[1] A comparison of the resulting subsamples with the full samples suggests that claims involving severe injuries or payment to the plaintiff are more likely to have good data and hence, are overrepresented in the analysis samples.

---

[1] The survey instrument, data collection and processing, response rates, and basic data are described in Westat (1978). Although 54 companies were writing over $1 million of malpractice insurance coverage in 1976, the nine participating companies accounted for 84 percent of all claims closed by private carriers in 1976.

[2] The survey instrument in all three years asked the number of defendants named in the case, but the surveys were not designed to include the claim files for all defendants in cases involving multiple defendants. Some cases required judgment to determine which claims referred to the same incident. Claim files for many defendants were missing, probably most commonly in incidents involving both a hospital and a physician, where the hospital self-insures or is insured by a different company. For example, in the 1976 survey, 7,759 defendants are reported to have been named in the 4,126 incidents, but survey forms were completed for only 4,812 of these defendants. Where multiple files for a single incident were present and differences existed in reported items that should have been identical, I used the value from the file for the defendant on behalf of whom the largest payment was made, on the presumption that the incentive to collect accurate information is greatest where liability is greatest.

[3] Insurance companies encourage insureds to report any incident that may lead to a claim. In the 1970 survey, 28 percent of all claims were reported by an insured. Preclaims were excluded from the 1976 survey.

Data from other sources on characteristics of the state in which the claim occurred were merged with the basic claim files.

All variables used in the analysis are described in Table 2. Means, by stage of disposition and outcome, are given in Tables 3, 4, and 5.

---

[1] Key variables for all samples were: size of payment, stage of disposition, dates of the incident, filing and disposition of the claim. In addition, for the 1974 and 1976 samples, which contain information on economic loss, a case was excluded if either (1) zero economic loss was reported or (2) (in 1974 only) future wage loss was zero, the injury was permanent, and the plaintiff was either a student or currently employed. The 1976 survey did not distinguish between incurred and future wage loss. I include a dummy variable if reported loss is under $100, which is suspiciously low. The large positive coefficient of the variable confirms that loss is underreported in these cases.

Table 2

DEFINITIONS AND AVAILABILITY OF VARIABLES
(* = variable present in claim files)

Variable	Definition	1970	1974	1976
Injury[a]				
PPD	Permanent partial disability	(b)	*	*
PTD	Permanent total disability	(b)	*	*
Death	Injury resulted in death	*	*	*
Induced	Injury induced by treatment	n.a.	n.a.	*
Incomplete cure	Incomplete cure	n.a.	n.a.	*
Lack of prevention	Injury caused by lack of preventive measures	n.a.	n.a.	*
Misdiagnosis	Misdiagnosis alleged	n.a.	*	n.a.
Consent	Lack of informed consent alleged	n.a.	*	n.a.
Res ipsa	1974: *Res ipsa* alleged 1976: Obvious error (see text)	n.a.	*	*
Plaintiff Characteristics				
Loss	Medical + wage loss + other expense, incurred and future ($\log_e$)	n.a.	*	*
Loss dummy	Dummy = 1 if loss < $100	n.a.	*	*
Life	Life expectancy = 70 - age at time of injury	*	*	Discreet
Life^2	Life squared	*	*	*
PPD × Life	Life expectancy, if permanent partial disability	*	*	*
PTD × Life	Life expectancy, if permanent total disability	*	*	*
Death × Life	Life expectancy, if plaintiff deceased	*	*	*
Income	Annual income if employed ($\log_e$) (may include non-wage income)	six categories	Wage income	Discreet
Not employed	Dummy = 1 if not employed, or employed and income unknown[c]			
Attorney	Plaintiff represented by attorney	*	n.a.	*
Defense Characteristics				
Defendants	Number of defendants named ($\log_e$)	*	*	*
Insurance	Limits of liability insurance, per incident, summed over defendants ($\log_e$)	Discreet	n.a.	n.a.
Physician(s)	Dummy = 1 if at least one physician defendant	*	*	Number of physicians
Expense	Expenditure by defense, legal fees, and other allocated costs summed over defendants ($\log_e$)[d]	*	*	n.a.
Liability	Dummy = 1 if defendant's insurer considered claim involved negligence[e]	*	n.a.	n.a.
Severity	Dummy = 1 if defendant's insurer considered injury was definitely or probably as severe as claimed	*	n.a.	n.a.
Case Characteristics				
Lag in reporting	Months from injury to filing claim ($\log_e$)	*	*	*
Lag in disposition	Months from filing to disposition ($\log_e$)	*	*	*
Trial	Dummy = 1 if case settled during trial	n.a.	n.a.	*
Jury	Dummy = 1 if jury verdict	n.a.	n.a.	*
Award	Total payment to plaintiff, by all defendants ($\log_e$)	*	*	*
Award error	Suspected reporting error[f]	*	n.a.	n.a.

Table 2--Continued

Variable	Definition	1970	1974	1976
State Characteristics				
Lawyers' income	Median income of lawyers in 1969[g]			
Court delay	Average months from service of answer to trial, for personal injury cases tried by jury in state courts[h]			
Number of claims	Number of malpractice claims closed[i]			
Locality abandoned	Locality rule expanded or rejected by 1970[j]			
Res 70	Res ipsa applied to medical malpractice by 1970[j]			
Comparative negligence	Comparative negligence adopted[k]			
Fee limit	Limit on contingent fees[m]			
Collateral source	Law reducing award by recovery from other sources[m]			
Award limit	Dollar limit on recovery[m]			
Periodic payment	Law authorizing periodic payment for future damages[m]			
Standard of care	Law defining standard of care[m]			
Win rate	Proportion of claims closed with payment to plaintiff 7/75 to 7/76[m]			
Scope of liability	Standard of care + Informed consent[m]			
Insurance share	Market share of largest firm writing malpractice insurance[n]			
Excess ratio	(Incurred loss $> \$25{,}000$ per claim) ÷ (Incurred loss $\leq \$25{,}000$ per claim, 1970-74)[p]			
Informed consent	Law restricting action based on lack of informed consent[m]			

[a] The omitted category includes all temporary and minor permanent injuries, such as scarring or loss of a finger or toe.

[b] The 1970 survey used a nine code severity scale. For comparability with the other two datasets, category 6 is labeled PPD. This includes such injuries as loss of one limb, eye, kidney, or lung. Categories 7 and 8 are combined and labeled PTD. It includes such injuries as loss of two limbs, blindness, quadraplegia, and severe brain damage.

[c] The response rate on plaintiff's income was 33 percent in 1970.

[d] This variable is biased down because some adjustment costs are not allocated and files for some defendants were missing from the survey.

[e] Dummy = 1 if a response of "definitely yes" or "probably yes" was given to either of the following:

1. The claim was legally meritorious in terms of liability, or
2. There was a medical injury, and
   There was negligence or breach of professional duty, and
   The medical injury was causally linked to negligence or breach of duty.

[f] Reported total award exceeds sum of reported payments by individual defendants. Because this variable did not affect coefficients in the 1970 estimates, it was not computed for 1974 and 1976.

[g] Excludes persons working less than 27 weeks or reporting income less than $1,000. Source: *1970 Census of Population.*

[h] Averaged over the major urban areas in each state. Source: *Calendar Status Study,* Institute of Judicial Administration.

[i] For 1970, number of claims closed in each state was estimated as the sum of the (inverse of the) weights of claims included in the analysis file. It may be biased to the extent the survey sample is non-random (because of underrepresentation of hospitals) or the subsample selected for analysis is non-random (because of non-random underreporting on key variables). For 1974, same as 1970. For 1976, number of claims closed between July 1975 and June 1976. (NAIC, 1977)

[j] Source: Dietz, Baird and Berul (1973).

[k] App. 135 *California Reporter* 497.

[m] NAIC, 1977. Dummy = 1 if law enacted by 12/75 and not overruled.

[n] Steves and McWhorter, 1975.

[p] ISO (1976).

Table 3

MEANS, BY STAGE OF DISPOSITION AND OUTCOME: 1970

(Unweighted)

Variable	Pre-Suit		Suit		Verdict	
	No Award	Award	No Award	Award	No Award	Award
PPD × Life	1.05	.92	2.64	4.44	2.42	5.42
PTD × Life	.29	.47	.69	2.43	.67	.89
Death × Life	4.56	2.75	5.74	6.77	5.31	3.16
Defendants	.22	.20	.59	.64	.45	.54
Severity	.36	.73	.33	.68	.42	.68
Attorney	.60	.62	.98	.99	.98	1.0
PPD	.04	.03	.08	.14	.09	.18
PTD	.01	.02	.02	.06	.01	.05
Death	.15	.08	.16	.20	.20	.08
Not employed	.90	.87	.86	.79	.80	.71
Liability	.12	.76	.11	.75	.06	.63
Award error	0.0	.08	0.0	.23	0.0	.21
Physician	.62	.59	.69	.76	.78	.68
Locality abandoned	.73	.69	.74	.79	.66	.81
Defendants	.22	1.38	.59	.64	.45	.54
Life	31.11	30.85	30.31	32.43	27.73	25.21
Award	0.0	7.12	0.0	8.86	0.0	9.98
Insurance	11.86	11.90	12.10	12.09	12.12	12.04
Expense	2.20	2.97	6.33	7.16	7.81	8.74
Income	.86	1.18	1.29	1.86	1.75	2.59
Life^2	1405.	1402.	1210.	1367.	1036.	858.
Lawyers' income	9.92	9.93	9.94	9.96	9.91	9.94
Lag in reporting	1.63	1.34	2.63	2.17	2.37	2.52
Lag in disposition	2.05	1.71	3.02	3.32	3.09	3.56
Court delay	2.93	2.96	2.98	3.08	2.91	2.94
Number of claims	5.43	5.41	5.76	5.91	5.30	5.69
n (unweighted)	400	465	216	553	86	38

Table 4

MEANS, BY STAGE OF DISPOSITION AND OUTCOME: 1974

Variable	Pre-Suit		Suit		Verdict	
	No Award	Award	No Award	Award	No Award	Award
Lag in disposition	1.99	1.64	2.87	3.15	3.27	3.28
Life	29.10	29.29	26.68	28.82	26.82	29.61
Misdiagnosis	.22	.19	.33	.35	.43	.36
Consent	.05	.08	.18	.21	.37	.18
Res ipsa	.02	.07	.08	.21	.18	.21
PPD	.13	.18	.29	.34	.33	.28
PTD	.03	.03	.01	.06	.07	.09
Death	.14	.04	.12	.17	.13	.25
Physician	.66	.57	.82	.85	.83	.79
Life^2	1376.	1358.	1014.	1169.	997.	1183.
PPD × Life	2.97	5.10	6.81	9.67	7.52	10.35
PTD × Life	.51	.11	.16	1.25	1.23	2.89
Death × Life	3.10	.68	2.53	4.62	3.91	7.08
Defendants	.07	.10	.14	.33	.16	.44
Income	1.77	2.21	3.49	3.95	3.41	4.04
Loss dummy	.19	.15	.09	.02	.07	.04
Loss	6.54	6.77	7.68	8.52	8.19	8.64
Award	0.0	7.52	0.0	9.23	0.0	10.11
Locality abandoned	.62	.78	.68	.83	.54	.67
Lawyers' income	9.87	9.91	9.91	9.96	9.89	9.92
Court delay	2.78	2.63	2.78	2.63	2.78	2.81
Comparative negligence	.45	.46	.38	.29	.33	.36
No fault	.31	.37	.34	.24	.29	.34
Expense	4.73	4.87	6.97	7.81	8.37	8.59
Not employed	.81	.75	.61	.56	.62	.55
Lag in reporting	1.19	1.22	2.56	2.28	2.31	2.09
Number of claims	6.31	6.53	6.60	7.15	6.37	6.64
n	727	375	323	979	147	67

Table 5

MEANS, BY STAGE OF DISPOSITION AND OUTCOME: 1976

	Pre-Suit		Suit		Verdict	
Variable	No Award	Award	No Award	Award	No Award	Award
Loss	3.02	5.02	4.71	5.73	6.78	6.24
PPD	.15	.14	.21	.26	.30	.37
PTD	.03	.02	.06	.06	.05	.09
Death	.12	.05	.17	.19	.20	.37
Induced	.45	.53	.51	.54	.56	.67
Lack prevention	.19	.23	.06	.13	.10	.05
Incomplete cure	.30	.17	.36	.28	.27	.23
Life	29.45	29.88	30.70	31.30	29.32	29.41
PPD × Life	4.21	3.78	5.98	8.19	8.67	9.88
PTD × Life	1.16	.58	1.85	2.25	2.03	4.52
Death × Life	3.84	1.27	4.94	6.32	5.55	11.28
Attorney	.56	.63	.97	1.00	.95	1.00
Trial	n.a.	n.a.	.05	.04	n.a.	n.a.
Jury verdict	n.a.	n.a.	n.a.	n.a.	.75	.88
Physicians	.70	.54	.85	.83	.99	1.33
Lag in reporting	1.55	1.38	2.71	2.34	2.47	2.26
Lag in disposition	1.95	1.54	2.82	3.14	3.20	3.59
Loss dummy	.58	.25	.40	.32	.20	.33
Fee limit	.26	.31	.29	.34	.16	.28
Collateral source	.28	.30	.31	.38	.26	.40
Award limit	.18	.19	.18	.20	.10	.12
Periodic payments	.24	.20	.19	.19	.15	.16
Standard of care	.14	.09	.09	.06	.17	.05
Comparative negligence	.51	.52	.50	.45	.54	.60
Lawyers' income	9.93	9.94	9.92	9.96	9.89	9.95
Court delay	2.81	2.82	2.84	2.92	2.83	3.06
Insurance share	62.87	62.46	63.45	62.02	66.65	65.30
Defendants	.18	.14	.64	.59	.59	.81
Award	0.0	7.53	0.0	9.12	0.0	10.76
Informed consent	.21	.19	.22	.27	.28	.30
Win rate	.37	.42	.41	.43	.37	.43
Excess Ratio	1.40	1.45	1.47	1.48	1.48	1.45
Life^2	1276.	1290.	1242.	1329.	1132.	1139.
Income	1.79	2.26	3.29	3.15	5.02	4.72
Number of claims	6.55	6.51	6.63	6.85	6.35	6.86
Not employed	.80	.75	.64	.66	.45	.49
n	895	486	548	1141	135	43

V. SIZE OF AWARDS

THE LEGAL STANDARD

In principle, the size of an award is determined by legally compensable damages suffered. For all injuries other than death, there are two components of compensable damages. "Economic loss" (referred to hereafter as pecuniary loss or simply loss) includes medical expense, wage loss, and other tangible out-of-pocket costs, incurred and expected for the duration of injury. If the plaintiff was not employed at the time of the injury, potential wages or the cost of hiring someone to perform the services performed by a homemaker must be estimated. If the award is to be paid in a lump sum, future costs are discounted to present value.[1] In addition, compensable damages include a "reasonable amount" for "pain and suffering," which is a catchall for damages that do not have a ready market value (referred to hereafter as non-pecuniary loss).

In cases involving death of the injured party, compensable damages are determined by the state's wrongful death statute. There is considerable diversity in what is compensable in wrongful death actions.[2] However, the different rules all imply that compensable damages are less if the plaintiff dies than if he suffers a severe injury but survives because the compensable loss to the decedent's survivors or estate excludes the value of his consumption and non-pecuniary loss had he lived.

The law of compensable damages differs across states in certain respects. Some courts have ruled that a per diem or mathematical formula cannot be used to compute pain and suffering.[3] Others have ruled that the jury cannot be instructed that awards are tax exempt. Since

---

[1] Lump sum payments were the norm until 1974. Some states have since made periodic payments for permanent injuries either mandatory or at the discretion of the judge.

[2] Wrongful death statutes are described in "Comment," *North Carolina Law Review*, Vol. 44, 1966, pp. 402-441.

[3] Dobbs (1973), p. 545.

1975, many states have enacted laws designed to hold down awards specifically for malpractice cases. These include ceilings on awards *in toto* or on the pain and suffering component (11 states), periodic rather than lump sum payments (9 states), and modification of the collateral source rule (17 states).[1]

IMPLICATIONS OF THE MODEL

The assumption that the courts are susceptible to influence by the litigants implies that, in addition to compensable damages, awards depend on the expected payoff to and cost of litigation inputs. Characteristics of a case that affect compensable damages directly also affect the award indirectly, by affecting the payoff, hence the expenditure on litigation.[2] Characteristics of a case that affect the cost of litigation inputs affect the award only indirectly, through expenditure on litigation. Out-of-court settlements are affected not only by the expected court award (hence by compensable damages and litigation costs) but also by the parties' estimates of the probability that the plaintiff wins.

FINDINGS BY OTHERS

Several previous studies have examined the relation between awards and pecuniary loss as reported in closed claim files with a view to determining the adequacy, equity, and efficiency of compensation through the tort system.[3] The consensus is that the tort system tends to overcompensate cases involving low pecuniary loss and undercompensate cases involving large pecuniary loss.[1] This is an important finding in view of the trend of recent legislation to limit large awards. Because pecuniary loss excludes all non-pecuniary damages, which probably increase at least in proportion to pecuniary loss, and awards include the plaintiff's legal costs, the implication is that compensation is even more regressive relative to full loss.

---

[1] NAIC (1977) summarizes recent legislation. Under the collateral source rule, compensation of the plaintiff's losses from other sources, such as private health insurance, is not relevant to the tort award.

[2] The net indirect effect is positive (negative) if plaintiff expenditure is more (less) effective than defense expenditure in influencing the size of award. See the appendix. Although the net effect of litigant expenditure is of major policy interest, it cannot be measured because data on plaintiff expenditure are lacking.

[3] For medical malpractice, see NAIC (1977), Westat (1978), and AIA (1976); for automobile injuries see DOT (1971); for product liability see ISO (1976).

These studies have cautioned that this finding may be erroneous and attributable to errors in the data on loss. Estimates of loss are often not reported in the claim files.[2] Where loss is reported, difference in the rate of discount of future expenses could lead to the observed regressive relationship.

Aside from reporting errors, it was argued above that theoretical considerations do not predict a strict correspondence between award and loss. Verdicts will be influenced by litigation inputs. Settlements will be influenced by both litigation inputs and estimates of the probability of proving negligence (q). If there is a systematic relation between q and loss (for example, the larger the loss the lower the q for which it is worth filing a claim), then if settlements approximate qV, settlements will appear to be regressive with respect to loss. A regressive relation would also be implied if the cost of going to court increases with loss for the plaintiff relative to the defense.

EMPIRICAL ESTIMATES

Separate equations are estimated for claims closed at verdict, settled at suit, and settled before suit, because coefficients are expected to differ by stage of disposition. A log transformation of the dependent and continuous explanatory variables is used because the distribution of awards in absolute dollar amounts is highly skewed. Analysis based on the raw data would yield estimates dominated by the few largest cases and would provide little information about the majority of cases. The log transformation reduces the influence of the largest cases on the estimates. It presupposes, quite plausibly, that stochastic errors are proportional to the size of the loss and that binary variables--such as the sex of the plaintiff or an allegation of a particular doctrine in a case--shift the relation between award and loss by a uniform percentage for all cases.[1] The estimates for 1976 are reported in Table 6. The estimates for 1974 and 1970 are in Tables 7 and 8. In the verdict equations, coefficient estimates and significance levels tend to be very sensitive to the specification. It is not possible to distinguish to what extent this is caused by juries' unpredictable behavior, "selection" of atypical cases into the verdict subset (because predictable cases are more likely to settle), or simply small sample size.[2]

---

[1] For example, AIA (1974) finds awards greater (less) than loss for cases with loss under (over) $25-50,000; NAIC finds awards greater (less) than loss for cases with loss up to (over) $70-79,999.

[2] In the 1976 survey, 29 percent of claims that reported injuries at least as severe as "temporary minor" were missing data on incurred medical expense.

Compensable Damages

The first regression for each year includes only a four part severity index, to show the gross variation in size of award by severity of the injury and stage of disposition. The intercept represents temporary and permanent minor injuries. In 1974 and 1976, for which the severity categories are more precisely defined, the severity index accounts for 13 to 20 percent of the variance in (log of) awards at pre-suit, roughly 25 percent at suit, and over 40 percent at verdict. Awards are therefore far from random with respect to severity of injury. However, within the category of temporary and permanent minor injuries, which account for over 50 percent of all claims, gradations

---

[1] A linear specification would imply that a random error of $x is equally likely for a minor injury or a severe injury and that a binary variable changes the expected award by $y, regardless of the severity of the injury.

[2] To test the importance of small sample size, a sample was selected from cases settled at suit that matched the jury sample on state and severity characteristics. Coefficients were as erratic and significance levels as low as with the verdict sample. Thus, the hypothesis cannot be rejected that the instability of coefficients in the verdict equations is due to the smaller sample of verdicts, rather than to an inherently less predictable process.

Table 6

AWARDS ($\log_e$), BY STAGE OF DISPOSITION: 1976

(t-statistics in parentheses)

Variable	Pre-Suit			Suit			Verdict		
Compensable Damages									
PPO	1.29 (6.05)	.54 (2.11)	.48 (1.97)	1.17 (11.54)	.31 (1.72)	.29 (1.63)	1.09 (1.96)	.85 (1.33)	1.32 (2.09)
PTD	1.91 (3.87)	-1.16 (-.24)	-.21 (-.43)	2.61 (13.97)	.24 (.73)	.14 (.42)	3.42 (4.44)	3.56 (3.26)	4.14 (3.45)
Death	1.95 (5.84)	-.008 (-.02)	-.001 (-.003)	1.13 (13.35)	.90 (4.83)	.93 (4.99)	2.38 (4.26)	2.71 (3.67)	3.06 (4.12)
Loss		.68 (18.58)	.68 (18.47)		.37 (12.91)	.37 (12.76)		.09 (.63)	.06 (.38)
Loss dummy		4.67 (16.66)	4.67 (16.71)		2.83 (11.26)	2.77 (11.10)		.36 (.26)	-.14 (-.10)
Life		-.004 (-.46)	-.003 (-.39)		.002 (.28)	.003 (.49)		-.08 (-1.54)	-.08 (-1.55)
Life^2		-.0000 (-.05)	-.0000 (-.14)		-.0000 (-.45)	-.0000 (-.68)		.0008 (1.22)	.0008 (1.20)
PPO × Life		.005 (.64)	.005 (.65)		.02 (3.64)	.02 (3.58)			
PTD × Life		.01 (1.02)	.02 (1.14)		.04 (4.82)	.04 (4.95)			
Death × Life		.03 (2.91)	.03 (3.00)		.007 (1.37)	.007 (1.35)			
Collateral source		-.02 (-.16)			.22 (2.41)			.37 (.75)	
Award limit		.02 (.10)			.07 (.53)			.24 (.27)	
Periodic payment		-.15 (-.46)			-.27 (-1.42)			1.04 (.95)	
Evidence of Negligence									
Induced		.20 (1.93)	.19 (1.86)		.24 (3.27)	.24 (3.24)			
Attorney[a]/Trial[b]/ Jury verdict[c]		.90 (7.4)	.87 (7.13)		.72 (3.79)	.68 (3.60)		1.12 (1.54)	1.87 (1.88)
Standard of care			-.05 (-.26)			-.07 (-.36)			
Win rate			.98 (1.57)			-.76 (-1.43)			-3.04 (-.88)
Litigation Costs									
Income		.08 (.53)	.08 (.54)		.23 (2.43)	.23 (2.44)		.68 (1.27)	.66 (1.27)
Not employed		.82 (.58)	.84 (.59)		2.19 (2.48)	2.19 (2.49)		6.76 (1.37)	6.69 (1.39)
Lag in reporting		.08 (1.75)	.08 (1.69)		-.09 (-2.92)	-.11 (-3.23)		.44 (2.39)	.41 (2.33)
Defendants		.24 (1.34)	.24 (1.34)		.19 (2.81)	.21 (3.20)		-.49 (-1.40)	-.61 (-1.64)
Physicians		.14 (1.35)	.14 (1.42)		.26 (5.25)	.24 (4.90)			
Court delay			-.16 (-1.10)			.20 (2.06)			
Lawyers' income			-.14 (-.38)			1.12 (3.22)			-.64 (-.32)
Comparative negligence			-.11 (-.90)			.21 (2.20)			-1.00 (-1.32)
Fee limit		.05 (.28)	-.18 (-1.15)		-.26 (-2.30)	.05 (.51)		-.74 (-1.26)	-.83 (-1.12)
Lag in disposition		.16 (2.97)	.17 (3.17)		.06 (1.20)	.04 (.73)		-.003 (-.009)	.10 (.28)
C	7.20 (87.14)	.97 (.67)	2.53 (.62)	8.37 (139.60)	2.91 (3.14)	-8.47 (-2.39)	9.15 (19.68)	1.59 (.28)	9.44 (.49)
R^2	.1352	.637	.640	.238	.453	.460	.433	.659	.6617
n	486			1141			43		

[a] Pre-suit only.

[b] Suit only.

[c] Verdict only.

Table 7

AWARDS ($\log_e$), BY STAGE OF DISPOSITION: 1974
(t-statistics in parentheses)

Variable	Pre-Suit			Suit			Verdict		
Compensable Damages									
PPD	1.47	.27	.32	1.26	.32	.33	1.79	.43	.46
	(6.76)	(1.10)	(1.33)	(12.50)	(1.83)	(1.94)	(3.57)	(.44)	(.55)
PTD	1.90	.53	.46	2.04	.55	.44	4.34	1.08	.13
	(3.89)	(1.34)	(1.19)	(10.61)	(1.94)	(1.58)	(5.80)	(.71)	(.10)
Death	2.19	.77	.75	1.76	.93	.87	2.22	1.14	.98
	(5.36)	(2.18)	(2.13)	(14.00)	(4.61)	(4.44)	(4.30)	(1.28)	(1.32)
Loss		.70	.68		.29	.26		.49	.37
		(16.53)	(15.97)		(10.79)	(9.74)		(4.28)	(3.82)
Loss dummy		-.43	-.38		-.34	-.29		-.68	-.76
		(-2.47)	(-2.22)		(-1.22)	(-1.09)		(-.74)	(-1.00)
Life		.01	.01		.02	.01		-.02	-.03
		(1.05)	(1.15)		(2.27)	(1.68)		(-.49)	(-.89)
Life^2		-.0002	-.0002		-.0003	-.0002		.0002	.0003
		(-1.21)	(-1.30)		(-2.88)	(-2.08)		(.29)	(.60)
PPD × Life		.001	-.002		.02	.01		.007	-.003
		(.20)	(-.29)		(3.27)	(2.67)		(.25)	(-.13)
PTD × Life		.03	.04		.02	.02		.02	.03
		(.59)	(.73)		(2.29)	(1.98)		(.53)	(1.01)
Death × Life		-.03	-.03		.005	.002		.004	-.004
		(-1.84)	(-1.9)		(.84)	(.43)		(.16)	(-.19)
Litigation Costs									
Income		-.03	.02		-.02	-.05		-.46	-.16
		(-.17)	(.12)		(-.18)	(-.46)		(-.86)	(-.36)
Not employed		.03	.45		.11	-.16		-4.17	-1.70
		(.02)	(.31)		(.12)	(-.18)		(-.86)	(-.43)
Lag in reporting		.14	.13		-.06	-.09		.30	.25
		(2.64)	(2.48)		(-1.5)	(-2.42)		(2.02)	(1.85)
Defendants		.52	.42		.57	.47		.30	.10
		(2.32)	(1.89)		(6.58)	(5.52)		(.67)	(.28)
Physician		.10	.11		.14	-.03			
		(.83)	(.88)		(1.20)	(-.31)			
Court delay			-2.18			.14			
			(-1.27)			(1.01)			
Lawyers' income			.38			.10			2.07
			(1.04)			(.27)			(1.31)
Comparative negligence			.10			.22			.46
			(.87)			(2.34)			(1.25)
Lag in disposition			.16			-.09			.09
			(2.41)			(-1.51)			(.42)
Expense			.19			.28			.60
			(2.41)			(7.84)			(4.32)
C	7.12	2.47	-2.25	8.39	5.93	3.52	8.65	9.01	-18.18
	(75.13)	(1.66)	(.56)	(126.2)	(6.38)	(.91)	(26.29)	(1.84)	(-1.13)
R^2	.173	.669	.687	.241	.408	.449	.403	.659	.798
n	375			979			67		

Table 8

AWARDS ($\log_e$), BY STAGE OF DISPOSITION: 1970
(t-statistics in parentheses)

Variable	Pre-Suit			Suit			Verdict		
Compensable Damages									
PPD	2.74 (5.68)	1.92 (2.64)	1.07 (1.63)	1.37 (7.53)	.39 (1.37)	.17 (.61)	.74 (1.33)	2.07 (1.91)	1.18 (1.15)
PTD	3.53 (5.03)	3.17 (3.13)	2.94 (3.23)	2.57 (8.44)	1.05 (1.64)	.78 (1.25)	1.18 (1.27)	5.93 (1.13)	5.46 (1.11)
Death	2.54 (7.75)	1.04 (2.07)	.76 (1.69)	1.21 (7.87)	.64 (2.42)	.55 (2.11)	.62 (.65)	.13 (.09)	-.20 (-.15)
Life		-.03 (-2.52)	-.03 (-2.75)		.02 (2.37)	.02 (1.83)		-.02 (-.35)	-.009 (-.25)
Life^2		.0003 (1.63)	.0003 (1.98)		-.0003 (-2.45)	-.0003 (-2.09)		.001 (1.17)	.0005 (.46)
PPD × Life		-.004 (-.27)	.003 (.20)		.02 (2.12)	.02 (2.54)		-.005 (-1.29)	-.03 (-.85)
PTD × Life		-.03 (-1.18)	-.05 (-1.90)		.02 (1.36)	.02 (1.45)		-.43 (-1.26)	-.38 (-1.20)
Death × Life		.02 (1.35)	.01 (1.33)		-.0009 (-.12)	-.001 (-.16)		-.005 (-.91)	.01 (.24)
Severity		.30 (1.91)	.16 (1.13)		.59 (5.30)	.62 (5.76)			
Evidence of Negligence									
Liability		.78 (4.89)	.83 (5.86)		.85 (7.66)	.84 (7.87)			
Attorney		1.48 (10.27)	.93 (6.62)						
Litigation Costs, etc.									
Income		-.31 (-.71)	-.11 (-.28)		.78 (3.12)	.71 (2.94)		2.77 (2.31)	1.83 (1.55)
Not employed		-3.56 (-.93)	-1.47 (-.43)		6.11 (2.73)	5.59 (2.57)		2.45 (2.30)	1.63 (1.55)
Lag in reporting		.19 (2.99)	.17 (2.91)		-.01 (-.25)	-.009 (-.19)		-.007 (-.03)	-.17 (-.71)
Defendants		.26 (2.39)	.15 (1.55)		.05 (1.38)	.04 (1.17)		.22 (.85)	-.17 (-.63)
Physician		.17 (1.17)	.16 (1.19)		.70 (5.69)	.55 (4.56)			
Insurance		-.07 (-.99)	.09 (-1.4)		.14 (2.64)	.14 (2.73)		.91 (3.19)	.55 (1.88)
Court delay			-.16 (-1.17)			.01 (.11)			
Lawyers' income			-1.40 (-3.24)			-.81 (-1.81)			-.16 (-.10)
Lag in disposition			.34 (5.97)			.08 (1.32)			-.07 (-.19)
Expense			.13 (5.51)			.20 (5.85)			.99 (2.83)
Award error		.82 (2.41)	.60 (1.95)		.82 (5.56)	.77 (5.35)		1.32 (1.58)	1.43 (1.89)
C	6.57 (81.94)	9.07 (2.29)	21.2 (3.99)	8.05 (113.9)	-1.69 (.71)	5.45 (1.07)	9.22 (34.91)	-26.95 (-2.43)	-19.45 (-1.19)
R^2	.194	.482	.591	.218	.485	.528	.085	.598	.720
n	465			553			38		

of severity (as reported in the severity code used in these surveys) had no additional explanatory power.

The median value of awards implied by these estimates, by severity and stage of disposition, is reported in Table 9.[1] At each stage of disposition, awards increase with the severity of the injury, except that cases involving death receive substantially less than cases involving total disabilities, as implied by the law defining compensable damages. For each level of severity, awards increase with stage of disposition, with verdicts considerably in excess of settlements. This is consistent with the hypothesis that settlements approximate the expected court verdict, discounted by the probability that the plaintiff would win.[2]

Comparisons of award levels over time are undermined by differences in the definition of injury categories between 1970 and the two later years, the change in sample (only 1970 approximates a national random sample), and few observations. In a comparison of the awards for death, where the injury is unambiguous, there is a clear upward trend between 1970 and 1974 but no systematic trend between 1974 and 1976, with verdicts increasing but settlements lower in 1976. Westat (1979) compared

---

[1] Because the regression is in logs, the coefficients imply predicted values of the median for each cell. This is a more representative measure of central tendency than the arithmetic mean for a distribution that is as skewed as is the distribution of awards. Estimates of arithmetic means can be obtained by correcting for the variance:

$$\hat{\bar{X}}_i = e^{\hat{M}_i} + \hat{\sigma}^2/2,$$

where $\hat{\sigma}^2$ is the sum of squared residuals.

[2] It is often alleged that there is an inverse relation between q and V--the larger the stakes the lower the probability of winning necessary to make it worthwhile to file a case. At the limit, this would imply no relation between settlement size and severity of the injury, if settlements approximate qV. The evidence clearly refutes this extreme view. But the finding that the ratio of verdicts to settlements is lower for permanent total disabilities is consistent with an inverse relation between q and V, but could be due to other factors. Necessary conditions for the relation S = qV are: (1) both parties have unbiased expectations of the outcome in court, are risk neutral, and have equal incremental litigation expense; (2) settlement occurs at the midpoint between the ask and the offer.

Table 9

MEAN AWARD, BY STAGE OF DISPOSITION AND SEVERITY OF INJURY: 1970, 1974, 1976[a]

	Temporary or Permanent Minor Disability		Permanent Partial Disability		Permanent Total Disability		Death	
Stage of Disposition	$	Sample Size	$	Sample Size	$	Sample Size	$	Sample Size
1970								
Pre-suit	713	405	11,048	14	24,343	9	9,045	37
Suit	3,134	332	12,332	77	40,946	33	10,509	111
Verdict	10,097	26	21,163	7	32,860	2	18,770	3
1974								
Pre-suit	1,236	281	5,378	68	8,267	11	11,048	15
Suit	4,403	421	15,522	333	33,860	59	25,591	166
Verdict	5,710	25	34,201	19	438,011	6	52,575	17
1976								
Pre-suit	1,339	384	4,866	68	9,045	10	9,414	24
Suit	4,316	673	13,905	297	58,688	68	13,360	103
Verdict	9,414	7	22,026	16	331,042	4	141,492	16

[a] Geometric means.

the simple arithmetic average award in 1976 and 1970 and found a 94 percent increase. If one decomposes the sample by type of injury and stage of disposition and focuses on median rather than mean values, this increase in the global average is apparently dominated by the increase in awards (especially verdicts) for the most severe injuries. Conclusions are tentative, however, because of small sample sizes for severe injuries.

These data cannot measure the adequacy of compensation. However, they can be used to investigate the related issue of whether awards increase more or less than in proportion to loss and the contribution to the observed relationship of reporting error, discounting of future loss, influence on settlements of q, and litigation costs.

First, to test whether the overcompensation of small losses and undercompensation of large losses reported by others can be attributed to measurement error in Loss, I estimated a bivariate regression between Award and Loss, in logs, in both directions.[1] For 1974, the estimated range for the true coefficient was: verdict: .83 to 1.48; suit: .62 to 2.33; pre-suit: .80 to 1.26. Thus the hypothesis cannot be rejected that the true coefficient is greater than or equal to unity--awards increase at least in proportion to loss and any conclusion to the contrary is attributable to reporting error in loss.

Second, if the estimated relation between award and loss were biased downward by failure to discount reported future loss or by influence on settlements of q and litigation costs, then the estimated coefficient of Loss should increase when proxies for these variables are included in the regression. In fact, although these variables have the signs predicted by this theory of settlements, the estimated coefficient of Loss does not increase significantly when they are added. This analysis is performed on a sample of cases excluding those where Loss is either zero or not reported or where specific components that should have been present were missing. The implication is that loss, as reported in insurance claims files, should not be used as an indicator of the adequacy of compensation, even for the cases where it is reported.[2]

When the other exogenous case characteristics are added in the second equation for each year, the percentage of the variation explained increases to over 60 percent at pre-suit, over 40 percent at suit, and over 60 percent at verdict. In the third equation, the characteristics of the state are added. They have little explanatory power, either individually or as a group.[1] Thus, awards are predictably related to the characteristics of the case. If there are systematic differences across states (as some evidence discussed below suggests) they are not explained by the attributes I have been able to measure.

---

[1] Because measurement error in the dependent variable does not affect the coefficient but measurement error in the independent variable biases the coefficient toward zero, running the regression in both directions yields bounds on the true value of the coefficient when both variables are measured with error.

[2] I tested whether the estimated coefficient of Loss differs between incurred and future loss, for the 1974 data where this distinction is made, on the presumption that incurred loss would be known with greater accuracy and discounting is not an issue. There was no significant difference in the estimated coefficients, suggesting that even incurred loss contains significant error. I also tested for significant difference in the coefficient of loss by severity of the injury. There was no significant increase in explanatory power over the specification reported here, which allows only the intercept to vary with severity.

The effect of plaintiff's life expectancy on size of awards is quite stable in all three years and consistent with expectations. With the 1974 estimates for claims closed at suit, awards first rise then diminish with age, peaking at age 37 for temporary and minor permanent injuries, which is consistent with age being an indicator of earnings and therefore loss. For more serious permanent injuries, awards are an additional 2 percent higher for each year of life expectancy, as expected if reported loss understates "true" loss more, the greater the severity and duration of the injury.[2]

---

[1] In 1976, some of the malpractice-specific state laws are included in the second regression. Estimates not reported here show that these variables have a negligible effect on the R^2 or the coefficients of the case-specific variables.

[2] Age is a categorical variable in the 1976 survey. In 1970, the coefficient of age is expected to differ because there is no measure of loss. Under certain assumptions, loss can be approximated by a log linear function of earnings at the time of the injury, severity of the injury, a quadratic of life expectancy and other variables:

$$L_{ij} = \int_0^D \alpha_i Y_{j0}^{\gamma} e^{\beta G_j} e^{[(\gamma\theta_1 - r)t - \gamma\theta_2 t^2]} dt,$$

where L_{ij} = present value of loss for individual j with injury level i
Y_0 = wage at time of injury
G = other characteristics
θ = growth rate of earnings
r = discount rate.

The assumptions necessary for this result are: (1) potential wages are a quadratic function of age; (2) loss is a uniform percentage, α_i, of potential wages for all individuals with an injury of level i; (3) other variables in G shift loss by a uniform proportion for all injuries. The finding that the coefficient of life expectancy changes little when Loss is omitted is further evidence of the error in reported loss.

The elasticity of award with respect to plaintiff's income is positive and greater at verdict than at suit in 1970 and 1976.[1] This is consistent with the hypothesis that income affects both compensable damages and the plaintiff's cost of litigating, and that the latter negative effect on the ask reduces the size of settlements, relative to verdicts.[2] There is no evidence that the absence of an observable measure of value of time results in lower wages for persons not employed in the market sector. Separate dummy variables were included for cases where the plaintiff was employed at the time of the injury but wage data are missing, and cases where the plaintiff was not employed. The coefficients were virtually identical. As further evidence, for the cases closed at suit in 1976, the predicted effect of income on award at the mean level of income for those reporting employment income is 2.13. The coefficient of the dummy for those not employed or with missing income data is 2.19. Finally, a dummy variable for the sex of the plaintiff was never significant.[3]

---

[1] The plaintiff's income from employment has the same direct and indirect effect on compensable damages as severity and duration of the injury. In addition, however, the opportunity cost of the plaintiff's time in litigation depends on his potential earnings. An increase in the cost of litigation inputs reduces the optimum amount purchased, other things equal, and increases total expenditure if the elasticity of expected payoff is less than unity. The net effect of income on the ask is ambiguous. It includes the positive direct effect on compensable damages and the negative time cost effect on litigation expenditure. By contrast an increase in plaintiff's income unambiguously increases the offer, because of the positive direct effect on compensable damages. Settlement size is therefore expected to increase with plaintiff's income unless the increase in the offer is more than offset by a decrease in the ask because of increased litigation costs.

[2] If Loss accurately incorporated the effect of income on compensable damages, Income would reflect only the partial effect of income on cost, so the expected sign would be negative. I tried to control for missing income data by including statewide averages: per capita income, mean manufacturing wage, percent of families in 1970 with income over $15,000, percent of persons over 25 with at least four years of high school. All were insignificant.

[3] Westat (1979) reports that the average non-zero award was $10,000 higher for male than female claimants. This is based on arithmetic averages, which are more sensitive to large cases than the log values used here and do not control for other characteristics of the case.

Evidence of Negligence

In principle, the size of award is unaffected by the degree of defendant negligence, in the absence of contributory negligence by the plaintiff or punitive damages for gross negligence. However, the model implies that settlements are positively related to the parties' expectations of the probability of a verdict for the plaintiff. To obtain a verdict of negligence, the plaintiff must prove that he incurred an injury, that the injury was causally linked to the medical treatment, and that the treatment fell short of the standard of care. Therefore, settlements depend on the cost of producing the evidence necessary to establish negligence. This depends on such factors as the type of injury, the quality of the evidence, the rules of evidence, etc., which are discussed further below.

The claim files contain very little information on the strength of the plaintiff's case. The evidence available, however, does support the hypothesis that size of settlements is positively related to the plaintiff's probability of winning. In 1970, the dummy variable, Liability, indicates if the insurance company believed the case was meritorious in terms of liability. The estimates imply that settlements are 80 percent higher where liability is acknowledged.[1] In 1976, the variable Induced indicates cases where the injury was reported to have been induced by treatment. Proof of negligence is expected to be relatively easy in these cases. (This is supported by evidence below.) Settlements are some 20 percent higher in these cases.

Finally, settlements at pre-suit are approximately 90 percent higher if the plaintiff is represented by an attorney.[2] This probability overstates the net contribution of the attorney to the size of award. Because plaintiff attorneys are typically paid a contingent fee, they have an incentive to reject cases with a low probability of winning. Therefore, representation is a rough indicator of the strength of the plaintiff's case.[1] The enactment of a law defining the standard of care, which was expected to affect q, had no discernible effect on the size of settlements in 1976.

---

[1] Settlements at suit are 60 percent higher where the insurer believed the injury was as severe as alleged. Both of these variables may be contaminated by hindsight of the survey respondents.

[2] Plaintiff representation is not reported for 1974. Representation is almost 100 percent for cases that proceed to suit, so this variable is included only in the pre-suit equations.

To the extent the data available do not permit accurate control for q in the size of settlement equations, coefficients of variables that are correlated with this omitted variable are biased.

Litigation Costs

The model implies that any factor raising the cost of litigation inputs tends to reduce the amount purchased and increase total expenditure if demand is inelastic, as expected.[2] An increase in the cost of litigation for the defense therefore implies an increase in the size of verdicts and of settlements (because of an increase in the offer); an increase in the cost of litigation for the plaintiff implies a decrease in the size of verdicts and of settlements (because of a decrease in the ask).

Two variables in the claims files are indicators of litigation costs for the defense. Defense costs increase with the number of defendants named. Because each defendant has an incentive to shift blame to the other, they are effectively working partly for the plaintiff, against each other.[3] The probability of a finding of negligence against at least one (a verdict for the plaintiff) is increased; the prediction with respect to the size of verdict is ambiguous, because all defendants share a common interest in holding down the size of the award. However, because settlements are positively related to the probability of a verdict for the plaintiff and to the litigation expense of the defense, settlements (as opposed to verdicts) are predicted to be positively related to the number of defendants.

---

[1] The Attorney coefficient incorporates at least three effects: (1) strength of the plaintiff's case; (2) size of the expected award--if there are fixed costs of bringing a case, an attorney will reject a case where his share of the expected payoff is less than this minimum cost (see Munch, 1980); (3) value added by the attorney--a plaintiff will hire an attorney only if he is expected to increase the payoff by more than his fee. The estimates imply that if imperfect control in our data for (1) and (2) accounts for less than a 57 percent differential between cases with and without representation, an attorney paid 33 percent more than earns his fee.

[2] See Munch (1980).

[3] In most states, the plaintiff controls the number of defendants named; named defendants cannot join unnamed defendants to the plaintiff's law suit. A named defendant can bring a separate suit for contribution or indemnity against a party who contributed to the injury but was not named by the plaintiff. The plaintiff's incentive is to name only those that are expected to increase his expected net payoff.

The marginal cost of additional defendants is expected to be highest in states that have adopted the doctrine of contribution in proportion to fault. Under this rule, each defendant can reduce his loss by shifting blame to the others. Under the alternative rule of pro rata (equal) contribution by all defendants found negligent, the payoff to shifting blame falls to zero for all defendants who expect to be found negligent, which facilitates cooperation among them to reduce the size of the verdict. As a proxy for contribution in proportion to fault, I include a dummy indicating state adoption of comparative negligence.

Litigation costs for the defense are higher for physicians than for institutional defendants, if the time and embarrassment cost of going to court is higher for physicians. This implies that settlements are positively related to the proportion of defendants who are physicians.

On the plaintiff's side, the positive effect of plaintiff income on costs has already been discussed. Two other variables in the claims files are indicators of plaintiff litigation cost. First, in 1970, the limits of the defendants' liability insurance coverage is reported. The higher the limits, the lower the cost of collecting a judgment, so the predicted relation between both verdicts and settlements and limits of coverage is positive. Second, the time elapsed between the occurrence of the injury and filing the claim is likely to increase the cost of marshalling evidence. This effect operates to a lesser degree

---

If expectations are correct, on average, a positive relation between payoff and number of defendants will be observed, not because there are unlimited increasing returns to naming additional defendants but because the unprofitable ones will not be observed.

for the defense because the burden of proof is on the plaintiff, except in *res ipsa* cases.[1]

Decay of evidence over time also implies that court delay increases the cost of going to trial for the plaintiff relative to the defense. This implies a reduction in ask relative to offer and, hence, a negative relation between court delay and the size of settlements.[2] However, court delay may also affect settlements in other ways. Because interest is generally not paid from the occurrence of the injury to the time of entering the judgment, delay reduces the present value of an award. This may be offset by monetary or "societal" inflation of awards over time or trends toward expanding the scope of liability, such that the probability of a finding for the plaintiff increases over time. If the parties have identical discount rates and expectations about the rate of inflation of awards, then their gains or losses from delay are offsetting and the net effect is analogous to a change in the expected verdict. If the expected rate of inflation is greater (less) than the rate of time discount, delay is equivalent to an increase (decrease) in the expected verdict. Assuming the net effect is negative, delay tends to reduce both parties' optimal expenditure on litigation and reduce the ask, the offer, and the size of settlement.[3] Because court delay affects so many factors, the net effect is unpredictable and may vary over time, between localities and between plaintiffs, depending on which of these several factors dominates.

---

[1] It has been alleged that the effective costs of going to court are higher for the plaintiff where the standard of care is that of the locality (a locality rule) because local physicians are unwilling to testify against each other. In an attempt to measure this "conspiracy of silence" I included the percent of practicing physicians in the state who are members of the state medical society, alone and interacted with the locality rule. Neither was significant.

[2] The average time between service of answer and trial, for personal injury cases tried by jury in state courts was 17 months in 1976. (Institute of Judicial Administration Calendar Status Study.) This figure is based on a sample of predominantly urban courts. Because delay is typically longer in urban than rural areas, the figure overstates the average over all courts but not necessarily the average relevant to malpractice claims. Munch (forthcoming) shows that malpractice claims per capita are more frequent in urban areas, largely because physicians are concentrated in urban areas.

[3] If the plaintiff's rate of discount or aversion to risk exceeds the defendant's, the ask falls by more than the offer and the size of settlement is reduced. The potential effects of court delay are discussed in Posner (1973).

Note: The discussion above presupposes that court delay imposes a constraint on the litigants. In fact, the average delay between service of answer and trial is less than the average time voluntarily chosen for negotiation.[1] A much more general approach to delay is to view time as an input in the dispute resolution process, analogous to attorney effort. For each party, time has a price that depends on expected nominal rate of inflation of awards, legal trends, rate of discount, and decay of evidence. Each party selects the optimum mix of delay and other litigation inputs, subject to the constraint of the court queue.[2] Delay will be substituted for dollar outlay, depending on their relative opportunity costs and marginal products in influencing outcome. For example, this model would predict that delay in disposition would be positively correlated with the wage rate of attorneys and with the potential wage income of the plaintiff. The productivity of delay relative to other legal inputs may be particularly high in the case of fairly severe injuries where only time can reveal the extent of the injury and narrow the range of expectations of possible outcomes.

An increase in the wage rate of attorneys decreases the optimal time input by both parties, but not necessarily by the same amount. The change in input in response to a change in attorney wage rates depends on the elasticity of expected payoff with respect to additional attorney time. If returns to attorney input diminish less rapidly for the plaintiff than for the defense, a decrease in attorney wage rates will induce a greater increase in input by plaintiffs than defendants, which tends to increase verdicts, increase the ask relative to the offer and increase the size of settlements. The opposite holds if the returns to additional attorney input diminish more rapidly for the plaintiff than the defense.[1] Ceilings on contingent fees for the plaintiff's attorney are expected to decrease the input of plaintiff attorney time and hence decrease awards and settlements, assuming that the ceilings are binding.[2]

---

[1] For example, in 1974, for the sample of cases settled with a positive award, mean time between filing and disposition of the claim is 28 months, whereas the mean court delay is only 14.8 months; in 1970 mean delay for cases settled is 34.1 months and mean court delay is 22.6 months.

[2] In a model of each party's optimum choice of delay, the assumption of independence is less plausible than in a model of the choice of other litigation inputs. Delay by one side automatically imposes a matching outlay by the other.

The evidence is generally consistent with predictions. Westat (1979) reports that on single defendant cases the average award is $12,218. The average time elapsed between filing and disposition is 15.2 months, and the average litigation expense $1,370. Comparable figures for multiple defendant cases are $46,275, 24.8 months, and $4,610.[3] However, if multiple defendant cases involve more severe injuries, the question arises how far these gross results reflect multiple defendants rather than severity. The multivariate analysis reported here shows that the positive relation between awards and number of defendants persists after severity of injury is controlled for, is greater at settlement than at verdict, and is greater for physician defendants. With the 1976 estimates, the elasticity of settlement size with respect to number of non-physician defendants (primarily hospitals) is .2, whereas for physicians it is .4, as predicted if physicians face a higher time and embarrassment cost of going to court.[1]

---

[1] It is plausible that returns to attorney input diminish less rapidly for the plaintiff than for the defense because the plaintiff's economic loss defines a floor to compensation, whereas the potential award for pain and suffering is limited only by the ingenuity of the plaintiff's lawyer. For some evidence in support of this and evidence that the elasticity of demand is probably less than unity, see Munch (1980).

[2] The types of ceilings on contingent fees for malpractice cases that have been enacted since 1975 are more likely to be binding on the few cases with large potential awards (severe injuries), particularly cases that are likely to be litigated. See Munch (1980) for estimates of the effects of binding fee constraints on size of awards.

[3] Many cases (43 percent) involved more than one defendant. The mean number of defendants per case was 1.94. Figures on litigation expense are from NAIC (1977). They are downward biased estimates of costs, because certain adjustment expenses are not routinely allocated to specific cases. They are additionally downward biased on multiple defendant cases because claim files for many defendants are missing.

The lag in reporting the claim, intended as a proxy for the cost to the plaintiff of proving his case, has the predicted negative sign for claims settled at suit, in 1974 and 1976. However, in both years it has a significant positive effect at pre-suit and verdict. This may reflect imperfect control for severity of the injury (the higher the costs, the greater expected award must be to make it worthwhile to file the claim) and non-random selection by stage of disposition and outcome (see below).

There is no evidence that court congestion induces malpractice claimants to settle for less, in contrast to the finding in Landes (1971) that the settlement rate in criminal cases is positively related to court delay. As noted above, court delay affects many aspects of the case in different ways. A negative net effect on settlement size is less likely if there is inflation of awards over time and if the delay in the courts is not a binding constraint on the parties' choice of bargaining time. Both conditions are more likely to apply to malpractice than to criminal cases.[2]

As predicted, awards are positively related to the limits of the defendants' insurance coverage, after number of defendants is controlled for. The estimated elasticity of award with respect to insurance limits is .9 at verdict, .14 at suit.[3] This suggests that the high limits of insurance coverage for medical malpractice compared with personal or automobile liability are one reason why awards for a given injury and venue tend to be higher for medical malpractice than for other areas of liability.

---

[1] The Physician variable is omitted from the verdict equations because the time and embarrassment cost of going to court is expected to raise the offer but not necessarily reduce the input of effort, given that a case goes to court. The coefficient of Physicians measures the differential effect per physician defendant in 1976. For the other two years, the dummy variable indicates the differential effect of at least one physician defendant.

[2] The difference between my results and those of Landes may also reflect correlation between court delay and other characteristics of the state. The coefficient is sensitive to which other state variables are included in the regression.

[3] It is insignificant at pre-suit, which is expected because the size of awards at pre-suit is typically well within the limits of the average policy.

Adding Defense Expenditure and Lag in Disposition has little effect on the other coefficients but significantly increases the R^2.[1] The highly significant, positive coefficient of Defense Expenditure is presumably due to correlation between this and other omitted variables--in particular, the plaintiff's litigation expenditure. More interesting is the finding that although cases settled before suit get higher awards the longer the time elapsed between filing and disposition, delay in disposition has no significant effect on award size for cases closed at suit or at verdict. The positive relation between size of award and lag in disposition reported in Westat (1979) is apparently due to the correlation between delay and other factors (in particular, severity of the injury). Delay *per se* has no effect, suggesting that awards do not include interest from the date of filing to closing, so the defense may gain by delaying settlement.

The only available measure of the wage rate of attorneys (median income in 1969) is not ideal. It has a significant negative effect on settlements in 1970, as expected if returns to attorney input diminish less rapidly for the plaintiff than the defense. However, this variable is positive but insignificant in 1974, positive and significant in 1976. It is impossible to tell whether this reflects a change in the judicial climate between 1970 and 1976, difference in the sample of states, larger measurement error in the variable, or correlation between this and other variables that changed over time. The enactment of a ceiling on plaintiff attorneys' fees has the predicted negative effect on verdicts and cases settled at suit when case characteristics only are included in the equation (1976 only). When other state variables are added (third equation), the ceiling on fees is no longer significant.

---

[1] Although these variables are endogenous, they are predetermined at the time the size of award is determined. There is insufficient information in the claims files to treat them as simultaneously determined with size of award.

Disposition

The 1976 estimates, if taken at face value, would imply that the 4 percent of cases settled during trial have 72 percent higher settlements, and that verdicts decided by a jury have awards more than twice as large as verdicts decided by a judge. No confidence should be placed on either of these estimates because of small sample sizes (only four cases decided by a judge) and possibility of selection bias.

VI. THE PROBABILITY THAT THE PLAINTIFF WINS

THE LEGAL STANDARD

In principle a medical provider is found liable if it can be shown that an injury occurred that was causally linked to the medical treatment, that the treatment fell short of the due standard of care, and the action lies within the scope of the provider's liability. In recent years changes in common law have expanded the scope of liability of hospitals and physicians. Abolition of charitable immunity and government immunity has brought non-profit and government hospitals within the scope of liability. The liability of hospitals for actions of their employees has been extended by the doctrine of *respondeat superior*. The doctrine of informed consent has made providers responsible for more than mere technical competence. Abolition of the locality rule has replaced a narrowly defined standard of care with a broader, often national standard. The doctrine of *res ipsa loquitur*, which shifts the burden of proof from the plaintiff to the defense, has made it easier for the plaintiff to get a case to the jury.[1]

Since the malpractice "crisis" of 1974-75, many states have enacted legislation attempting to limit the liability of providers. As of May 1977, 21 states had passed laws limiting the use of informed consent, 16 had modified the standard of care, and 16 had modified the standard of proof.

IMPLICATIONS OF THE MODEL

Rules that expand the scope of liability (increase in Z_2) increase the probability of a verdict for the plaintiff (q) directly. There is also an indirect effect through changes in the parties' expenditure on litigation. Expenditure by both parties increases, unless an increase in Z_2 reduces the marginal productivity of legal inputs in influencing q.[1] If the productivity of plaintiff inputs increases relative to defense, then the indirect effect reinforces the direct effect. An increase in the cost of litigation for the plaintiff (defense) tends to reduce (increase) q.

---

[1] The narrow interpretation of this doctrine applies to only a small number of cases of obvious error (such as the proverbial sponge in the stomach). The broader interpretation requires the plaintiff to prove only that it was more likely than not that the injury resulted from negligence, so is potentially applicable to many more cases.

For cases settled out of court, the probability that the plaintiff wins (q^*) was analyzed above in terms of the alternative probability that the case is dropped. A necessary condition is that the plaintiff's expected payoff from continuing exceeds his expected cost. The case will not be dropped if the defense actually makes a positive offer. The model implies, therefore, that the probability that the plaintiff wins, for cases settled out of court, is positively related to q, the expected award, and the litigation costs of the defense and negatively related to the litigation costs of the plaintiff.

The model outlined here does not formally include errors in expectations. Yet if expectations were always correct, cases would not be filed and then dropped, because this represents an investment with zero return. Cases where there is great uncertainty (large variance of expectations) as to the size of the award, the probability of winning, or the costs of litigation are more likely to be dropped.

EMPIRICAL ESTIMATES

Discriminant function estimates of the determinants of the probability of the plaintiff winning are reported in Tables 10-12 for 1976, 1974, and 1970.[2]

---

[1] It is conceivable that an increase in Z_2 increases the productivity of inframarginal inputs but reduces it at the margin. In this case, the effect of an increase in Z_2 on L is uncertain. See the appendix. Evidence from the 1970 survey shows that the defense spends more on cases it considers legally meritorious, which implies $\partial L^d/\partial q > 0$.

[2] If the underlying model of the probability of an event is logistic, discriminant function estimates are unconditional maximum likelihood estimates of the parameters of the logistic function, if the observations on the explanatory variables are drawn from a multivariate normal population. Given the normality assumption, t- and F-statistics computed by the discriminant function provide precise t- and F-tests for eliminating insignificant variables. When the normality assumption is not satisfied, as in the case of dummy variables, discriminant function estimates are not consistent. However, it has been shown by Halperin, Blackwelder, and Verter (1971) that discriminant function estimates of

Table 10

PROBABILITY THAT THE PLAINTIFF WINS: 1976

(t-statistics in parentheses)

Variable	Pre-Suit	Suit	Verdict
Compensable Damages			
Loss	-.11 (-2.09)	.20 (4.59)	.12 (.84)
Loss dummy	-2.33 (-5.91)	1.27 (3.36)	2.33 (1.78)
PPD × Life	.001 (.91)	.01 (3.09)	.04 (2.34)
PTD × Life	.004 (.37)	.007 (1.19)	.03 (1.27)
Death × Life	-.01 (-1.84)	.009 (2.43)	.04 (2.54)
Evidence of Negligence			
Res ipsa	1.71 (4.07)	1.08 (2.48)	-.64 (-.40)
Lack of prevention	.90 (2.84)	.88 (2.84)	-1.69 (-1.35)
Incomplete cure	-.03 (-.09)	.12 (.42)	-1.64 (-1.49)
Induced	.53 (1.88)	.46 (1.70)	-.35 (-.33)
Standard of Care	.12 (.53)	.07 (.29)	-.48 (-.60)
Informed consent	-.11 (-.68)	.28 (2.07)	-.77 (-1.37)
Win rate	5.21 (6.77)	2.79 (3.65)	9.53 (3.06)
Attorney[a]/Trial[b]/ Jury verdict[c]	.88 (5.86)	-.40 (-1.43)	.80 (1.26)
Litigation Costs, etc.			
Income	-.03 (-.12)	-.17 (-1.17)	-.20 (-.40)
Not employed	-.28 (-.13)	-1.36 (-.98)	-1.91 (-.41)
Lag in reporting	-.19 (-3.09)	-.25 (-4.57)	-.31 (-1.44)
Defendants	-.30 (-1.57)	-.28 (-2.94)	.54 (1.36)
Court delay	.60 (3.53)	.27 (2.07)	1.75 (2.81)
Lawyers' income	-.36 (-.70)	2.10 (4.17)	2.50 (1.18)
Comparative negligence	.01 (.09)	-.04 (-.32)	.10 (.19)
Lag in disposition	-.65 (-8.37)	.40 (5.04)	-.05 (-.15)
C	1.37 (.76)	-2.31 (-3.96)	-3.44 (-1.52)
R^2	.231	.123	.276
n	1181	1689	178

[a] Pre-suit only.

[b] Suit only.

[c] Verdict only.

Table 11

PROBABILITY THAT THE PLAINTIFF WINS: 1974

(t-statistics in parentheses)

Variable	Pre-Suit	Suit	Verdict
Compensable Damages			
Loss	.17	.18	-.07
	(.32)	(3.67)	(-.62)
Loss dummy	-.12	-.80	-.56
	(-.56)	(-1.99)	(-.67)
Life	.004	.008	-.008
	(1.02)	(1.83)	(-.61)
PPD × Life	.01	.006	.02
	(2.39)	(1.22)	(1.45)
PTD × Life	-.05	.02	.04
	(-2.38)	(1.47)	(1.29)
Death × Life	-.02	.0009	.03
	(-3.15)	(.13)	(1.71)
Evidence of Negligence			
Res ipsa	1.17	.73	.16
	(3.19)	(3.92)	(.34)
Informed consent	.37	-.08	-1.09
	(1.25)	(-.44)	(-2.78)
Misdiagnosis	-.24	.05	-.67
	(-1.36)	(.32)	(-1.77)
Locality	.64	.64	.82
	(2.99)	(2.37)	(1.44)
Litigation Costs, etc.			
Income	-.87	-.34	.07
	(-3.25)	(-1.81)	(.16)
Not employed	-8.03	-3.00	.22
	(-3.32)	(-1.80)	(.06)
Lag in reporting	.007	-.29	-.27
	(.10)	(-4.34)	(-1.60)
Defendants	.76	.67	1.65
	(2.51)	(4.10)	(3.72)
Court delay	-1.10	-.82	.95
	(-4.43)	(-3.04)	(1.51)
Lawyers' income	1.46	.74	1.29
	(2.29)	(.93)	(.74)
Comparative negligence	.10	-.07	-.09
	(.57)	(-.32)	(-.18)
Lag in disposition	-.71	.31	-.37
	(-7.69)	(3.16)	(-1.32)
Number of claims	-.23	.12	
	(-2.13)	(.10)	
Expense			2.11
			(1.06)
C	-3.23	-3.78	-.17
	(-.10)	(-.05)	(-.76)
R^2	.153	.147	.203
n	1102	1302	214

Table 12

PROBABILITY THAT THE PLAINTIFF WINS: 1970

(t-statistics in parentheses)

Variable	Pre-Suit	Suit	Verdict
Compensable Damages			
PPD	-1.63 (-1.53)	.33 (.53)	-1.84 (-1.00)
PTD	-.11 (-.08)	-1.26 (-1.07)	5.93 (1.72)
Death	-.38 (-.68)	.33 (.59)	-1.53 (-.94)
Life	.001 (.29)	.007 (.96)	-.03 (-1.46)
PPD × Life	.04 (1.66)	-.003 (-.15)	.09 (1.40)
PTD × Life	.01 (.33)	.02 (.84)	-.17 (-1.27)
Death × Life	-.006 (-.44)	-.01 (-.96)	.02 (.42)
Evidence of Negligence			
Liability	4.76 (24.48)	3.93 (18.94)	4.73 (5.89)
Attorney	.73 (3.42)		
Litigation Costs, etc.			
Income	.24 (.32)	-.68 (-1.22)	-.79 (-.59)
Not employed	1.98 (.30)	-6.66 (-1.32)	-7.40 (-.63)
Lag in reporting	-.14 (-1.71)	-.22 (-2.50)	.58 (2.03)
Defendants	-.07 (-.23)	-.08 (-.40)	-.77 (-1.08)
Court delay	.22 (.97)	.47 (1.61)	.19 (.28)
Lawyers' income	.51 (.62)	2.04 (2.02)	.92 (.43)
Lag in disposition	-.36 (-4.05)	.23 (1.93)	.56 (1.34)
Expense			.42 (1.39)
Number of claims	-.07 (-.60)	-.27 (-2.33)	
C	-8.84 (-.51)	-1.48 (-.96)	-10.48 (-.27)
R^2	.459	.366	.391
n	865	769	124

Evidence of Negligence

The claims files contain very little information on the evidence of negligence. The single most powerful predictor of payment being made to the plaintiff is the insurance company's evaluation that the claim had legal merit (Liability), which is reported only for 1970.[1] This indicates that the outcome in court is far from a random event, at least for the defense. Of cases won by the plaintiff, the insurer considered 51 percent had legal merit, but only 5 percent of cases lost by the plaintiff. Thus the outcome was correctly predicted 83 percent of the time.[2]

In 1974 it is known if a case alleged *res ipsa*. As expected, these cases have a significantly higher probability of receiving payment before verdict.[3] Similar effects are observed in 1976 for cases I classified as *res ipsa* cases because they involved an obvious error, such as the wrong patient or wrong drug. Cases involving an injury

---

parameter values and standard errors are quite close to the conditional maximum likelihood estimates, even when applied to data sets with several discrete variables. Because of the costs of using maximum likelihood techniques for exploratory analysis in large data sets such as these, only discriminant function estimates have been made so far. They are expected to provide reliable tests of significance but possibly biased estimates of parameter values.

[1] On the average defense expenditure on cases ultimately won by the defense was only 50 percent of the average amount spent on cases won by the plaintiff, which is highly correlated with Liability. This suggests that this variable is not mere hindsight of the survey respondent.

[2] The defense may choose to litigate cases it expects to lose for at least three reasons. First, on cases won by the plaintiff, the defense is unsure of the outcome: 51 percent accurate on average. Second, the size of the award is uncertain: in 63 percent of cases won by the plaintiff the defense considered the injury as severe as claimed. Because of some combination of these factors, settlement was recommended to the insured in only 36 percent of the cases ultimately won by the plaintiff. Third, the insured may refuse to settle. However, there are no reported cases where the insurer recommended settlement and insured approval was required but was refused. Thus there is no evidence that physicians' reluctance to implicitly admit liability by settling is a major obstacle to settlement. These conclusions are based on small samples, however. Percentages are lower bounds because nonresponses were categorized as negative responses.

[3] The apparent lack of effect at verdict is probably due to small sample size and selection bias: where negligence is obvious, litigation to verdict is less likely.

induced by treatment or alleging lack of preventive care are also more likely to receive payment than cases alleging incomplete cure or the omitted category, which includes misdiagnoses and unauthorized treatments. This is corroborated by the 1974 data, where cases alleging misdiagnosis are found to have a lower probability of winning at verdict, as are cases alleging lack of informed consent.[1] This evidence suggests that the likelihood of a finding of negligence is greater for obvious errors. If so, any deterrent effect of the malpractice system is uneven across medical procedures, depending on the patient's ability to detect and prove an error in treatment.

The probability of winning before suit is significantly higher if an attorney is present. Again, it is impossible to distinguish the marginal product of the attorney from the selection bias, because representation is less likely for cases with a low probability of winning.

In 1976 a variable indicating the average win rate by plaintiffs in the state over the year July 1975 to June 1976 is included. The significant positive sign at all stages of disposition confirms that in some states there is a higher propensity to hold for the plaintiff that is not captured by the variables I have been able to measure.

As in the case of awards, it was not possible to distinguish net effects of the 1975-76 malpractice reform legislation. Enactment of a law defining the standard of care was estimated to have a significant negative effect on the probability of winning, but this effect is reduced when Court Delay is added and eliminated when Lawyers' Income is added.

Compensable Damages

In principle, the severity of the injury is not relevant to the issue of negligence. The Learned Hand rule for determining negligence requires weighing expected damages relative to prevention costs. In medical malpractice this weighing of costs and benefits in each case is typically replaced by the standard of practice in the profession as the benchmark of negligence. However, in practice the issue may amount to determining whether an observed bad outcome is more probably the unlucky result of acceptable care or the result of negligent care. Other things equal, the more severe the injury, the more likely it was due to substandard care. Thus a positive association between severity of the injury and propensity to hold for the plaintiff would not necessarily indicate that the courts were applying a no-fault rather than a negligence standard.

---

[1] In 1974 12 percent of cases alleged *res ipsa*; in 1976 I classified only 2 percent as *res ipsa* cases. The finding that cases alleging lack of informed consent have a lower probability of winning is somewhat surprising, because this is often not the only allegation but additional to an allegation of negligent care. Thus, it adds another possible ground for finding negligence even if the defendant's practice is found to conform to the technical standard of care.

The evidence confirms that the probability that the plaintiff wins at verdict and suit is positively related to the severity and duration of the injury (statistically significant only in 1976). Economic loss is also highly significant at suit in 1974 and 1976. Thus, these results confirm the prediction that the propensity to drop is inversely related to the expected payoff. However, the results for settlements at pre-suit are not consistent. This may be attributable to selection effects: of cases involving a large loss, only those with weak evidence of negligence will close without filing suit.

The evidence clearly refutes the common allegation that insurance companies can be forced to pay on frivolous claims, in order to avoid costly litigation. Roughly 50 percent of claims closed at settlement (which constitute over 90 percent of all claims) receive no payment.[1]

Litigation Costs

Two proxies for the plaintiff's litigation costs are included: income and the lag in reporting the claim. The probability of dropping the case is positively related to the plaintiff's income but the significance level is low.[1] The lag in reporting the claim significantly increases the probability that a case is dropped.[2]

---

[1] It is true that claims receiving payment before suit tend to be low valued cases. But it has been argued that this results from selection of more severe cases out of this sample.

The effect of number of defendants (a proxy for the costs of the defense) is not consistent in all three years. In 1974 the probability of payment to the plaintiff is significantly positively related to the number of defendants, as predicted. However, in 1976 the effect is (insignificantly) positive at verdict but negative at settlement, and in 1970 it is insignificant. Contrary to expectations, the doctrine of comparative negligence has no (measurable) effect.

The effect of court delay on the probability that the plaintiff wins is also unstable over time--positive in 1976, negative in 1974, and insignificant in 1970. This may reflect real changes in the relative advantages of delay to the two parties, which is theoretically quite plausible but probably also reflects other unmeasured factors. To test this, Court Delay was included in the verdict equations. It has a significant positive effect in 1976 and 1974, which is not predicted by the theory that court delay operates by its effect on the expected payoff to the parties from litigating.

The wage rate of attorneys has either no effect or a positive effect on the probability that the plaintiff wins. This is inconsistent with the findings in the award equations. As in that case, the coefficients of the state variables are unstable.

---

[1] The predicted sign is negative only if Loss accurately measures the positive relation between income and payoff.

[2] In addition to increasing plaintiff costs (relative to the defense), delay in reporting a claim may increase uncertainty, which also increases the probability of dropping.

VII. THE DECISION TO LITIGATE

IMPLICATIONS OF THE MODEL

A necessary condition for continuing litigation is that the plaintiff's minimum ask exceeds the defendant's maximum offer; equivalently, the plaintiff's expected payoff exceeds the defendant's by more than the sum of their incremental litigation costs. It follows that the propensity to litigate is likely to be positively related to the stakes (V);[1] negatively related to plaintiff and defendant litigation costs; and positively (negatively) related to the probability that the plaintiff wins (q), if dispersion of expectations increases (decreases) with q.

EMPIRICAL ESTIMATES

Discriminant function estimates of the propensity to proceed from pre-suit to suit and from suit to verdict are reported in Tables 13-15 for 1976, 1974, and 1970. The data available in the claim files explain less of the decision to go to verdict than of the decision to file suit, which is consistent with the hypothesis that cases where the outcome is least predictable tend to go to verdict.

Compensable Damages

Cases involving relatively large loss are much more likely to file suit, but loss is unrelated to the decision to go to verdict. For minor injuries, the likelihood of early settlement is greater for young plaintiffs. However, for severe injuries, the likelihood of early settlement is less for young plaintiffs, as would be expected

---

[1] Landes (1971) and Posner (1973) conclude that the propensity to litigate increases with the stakes of the case. They reach this conclusion in the context of a model that does not explicitly allow for changes in parties' expenditure on litigation with the stakes, and the indirect effect of this change on the outcome. It is shown in the appendix that these indirect effects are offsetting, so the conclusions of the simple model hold.

Table 13

THE DECISION TO CONTINUE LITIGATION: 1976

(t-statistics in parentheses)

Variable	To Suit[a]	To Verdict[b]
Compensable Damages		
Loss	.28	.04
	(8.54)	(.68)
Loss dummy	1.44	.14
	(5.78)	(.25)
Life	-.009	-.01
	(-3.25)	(-1.95)
PPD × Life	.01	.01
	(4.12)	(1.83)
PTD × Life	.003	.01
	(.52)	(1.54)
Death × Life	.02	.01
	(5.08)	(2.40)
Limit on awards	-1.07	.50
	(-4.73)	(1.20)
Excess ratio	.46	-.09
	(3.69)	(-.40)
Evidence of Negligence		
Lack of prevention	.09	-.21
	(.63)	(-.75)
Incomplete cure	-.21	-.37
	(-2.04)	(-1.99)
Scope of liability	.009	.29
	(.12)	(2.30)
Litigation Costs, etc.		
Income	-.15	.20
	(-1.23)	(.96)
Not employed	-1.66	1.16
	(-1.45)	(.59)
Lag in reporting	.75	-.03
	(19.62)	(-.37)
Defendants	1.41	-.12
	(16.21)	(-.82)
Physicians	-.22	.30
	(-3.44)	(2.84)
Fee limit	.34	-.87
	(2.18)	(-2.95)
Court delay	.25	-.25
	(2.02)	(-1.09)
Lawyers' income	-.28	-1.95
	(-.68)	(-2.32)
Comparative negligence	-.20	-.38
	(-1.84)	(-1.71)
Insurance share	-.007	.02
	(-1.82)	(2.25)
Number of claims	.27	-.10
	(4.42)	(-.83)
C	-1.51	1.61
	(.36)	(2.43)
R^2	.310	.048
n	3248	1867

[a] 0 = claims closed pre-suit.
1 = claims closed at suit or verdict.

[b] 0 = claims closed at suit.
1 = claims closed at verdict.

Table 14

THE DECISION TO CONTINUE LITIGATION: 1974

(t-statistics in parentheses)

Variable	To Suit[a]	To Verdict[b]
Compensable Damages		
Loss	.31 (8.81)	-.02 (-.41)
Loss dummy	-.18 (-.95)	.74 (1.77)
Life	-.006 (-2.24)	-.008 (-1.63)
PPD × Life	.02 (4.87)	.004 (.71)
PTD × Life	-.003 (-.34)	.02 (1.79)
Death × Life	.02 (3.80)	.01 (1.44)
Evidence of Negligence		
Res ipsa	1.17 (7.45)	.15 (.75)
Misdiagnosis	.18 (1.55)	.40 (2.40)
Informed consent	.65 (4.43)	.78 (4.09)
Locality abandoned	.56 (3.34)	-.50 (-1.68)
Litigation Costs, etc.		
Income	-.54 (-3.64)	.05 (.23)
Not employed	-5.23 (-3.93)	.48 (.26)
Lag in reporting	.85 (19.10)	-.10 (-1.36)
Defendants	.61 (4.38)	-.18 (-.97)
Physician	.21 (1.63)	-.11 (-.51)
Court delay	.57 (3.16)	.45 (1.55)
Lawyers' income	.07 (.14)	-.74 (-.85)
Comparative negligence	-.34 (-2.38)	-.23 (-.96)
Number of claims	.21 (2.57)	-.22 (-1.67)
C	-2.78 (.09)	6.01 (1.08)
R^2	.368	.056
n	2618	1516

[a] 0 = claims closed pre-suit.
1 = claims closed at suit or verdict.

[b] 0 = claims closed at suit.
1 = claims closed at verdict.

Table 15

THE DECISION TO CONTINUE LITIGATION: 1970

(t-statistics in parentheses)

Variable	To Suit	To Verdict
Compensable Damages		
PPD	1.56	-.05
	(3.82)	(-.08)
PTD	.38	.02
	(.55)	(.02)
Death	.58	.05
	(1.89)	(.10)
Life	-.004	-.02
	(-1.11)	(-2.74)
PPD × Life	-.01	.01
	(-.98)	(.69)
PTD × Life	.02	.003
	(.94)	(.13)
Death × Life	.002	-.001
	(.30)	(-.09)
Severity	-.18	-.001
	(-1.50)	(-.01)
Evidence of Negligence		
Liability	.25	-1.50
	(2.07)	(-6.89)
Locality rejected	-.22	.13
	(-1.58)	(.47)
Litigation Costs, etc.		
Income	.64	-.91
	(1.89)	(-1.72)
Not employed	5.06	-8.70
	(1.67)	(-1.83)
Lag in reporting	.51	.09
	(10.72)	(.98)
Defendants	.91	.19
	(6.84)	(-.92)
Physician	.05	.31
	(.38)	(1.29)
Insurance	.05	.20
	(.85)	(1.83)
Court delay	-.02	-.26
	(-.16)	(-.94)
Lawyers' income	-.25	-1.49
	(-.52)	(-1.43)
Number of claims	.24	-.31
	(3.31)	(-2.40)
C	-5.90	2.13
	(-.60)	(2.30)
R^2	.190	.121
n	1758	893

if both the expectation and variance of the size of award increase with the severity and duration of the injury.[1]

Disagreement as to the severity of the damages is apparently a less important source of conflict than disagreement as to liability. The insurer's evaluation that the injury was as severe as claimed (Severity) is not significantly related to the propensity to litigate at either stage.

Evidence of Negligence

The most powerful predictor of whether a case will go to verdict is the insurer's opinion of the liability. Cases where the defense recognizes liability are likely to be settled.[2] Thus the fact that the plaintiff wins only 20 percent of the cases that go to verdict but 50 percent of the cases settled before verdict in part reflects the selection of cases with weak evidence of negligence into the verdict sample.

Cases alleging *res ipsa* in the 1974 sample are much more likely to proceed to suit, but not to verdict.[3] Cases alleging misdiagnosis or lack of informed consent are more likely to go to suit and to verdict, suggesting that plaintiffs systematically overestimate their chances of winning on these cases. Cases involving an incomplete cure are less likely to litigate and, as shown earlier, are less likely to win. Thus this type of case also appears to involve plaintiff overoptimism.

---

[1] In both 1976 and 1974, cases of permanent partial disability are more likely to be litigated to suit than cases of permanent total disability. It is plausible that uncertainty as to the extent and duration of the injury is greater in the case of partial than of total disability.

[2] The evidence suggests a negative correlation between q and variance of the parties' expectations of q. This assumes that the defense's prediction of q is unbiased. To the extent there is prediction error, cases where the defense overestimates q are more likely to settle, and the measured effect of Liability is an upward biased estimate of the effect of q.

[3] Cases classified as *res ipsa* in 1976 did not have a significantly different probability of being litigated, so this variable was dropped.

Litigation Costs

The propensity to file suit is negatively related to plaintiff's income in 1974 and 1976, which is consistent with the theory that an increase in income decreases the ask relative to the offer.[1] Contrary to predictions based on plaintiff costs, a long lag in reporting the claim increases the propensity to file suit but has no consistent effect on the propensity to go to verdict. The former effect may reflect an urgency to file suit before the running of the statute of limitations.

In 1976 the probability of going to verdict is significantly lower in states that had enacted a ceiling on contingent fees, although the propensity to file suit is higher in these states. As with all state variables, confidence in this conclusion is weak.

Cases involving multiple defendants are more likely to proceed to suit. The tendency of multiple defendants to raise costs for the defense (which should encourage early settlement) is apparently offset by greater uncertainty as to allocation of liability among the defendants. In 1976, I included a measure of the market share of the leading malpractice writer in the state, to test the hypothesis (not incorporated into the model) that the payoff to the defense from litigating one case has precedent value on other cases. The positive coefficient is consistent with this hypothesis.

There is no evidence that court delay encourages settlement. This variable is positively related to the decision to sue, unrelated to the decision to go to verdict. Again, this can be rationalized by correlation with unmeasured factors that encourage litigation in malpractice as well as other personal injury cases.

Other State Effects

It was hypothesized that uncertainty about court behavior should be less and settlement more common in states with a large number of claims. On the contrary, the number of claims in the state is positively correlated with the propensity to sue, presumably because both are correlated with unmeasured factors. The propensity to sue is also positively related to the excess limits ratio, which is an indicator of the chance of a large payoff.[1]

---

[1] The positive effect of income on compensable damages increases ask and offer symmetrically; the positive effect of income on litigation costs reduces the ask but does not affect the offer.

---

[1] The excess limits ratio is the ratio of the average incurred cost per claim in excess of $25,000 to the average incurred cost per claim truncated at $25,000, for 1970-74.

VIII. CONCLUSION

This analysis of the disposition of malpractice claims has used data from three surveys to address the following questions:

- To what extent does reality conform to the legal standard of payment equal to damages, if and only if negligence occurred?
- Are departures from the legal standard consistent with an economic model of claim disposition?
- What is the net effect of certain case characteristics on the outcome (as opposed to gross relationships tabulated in previous analyses of these claims files)?
- What is the effect of specific laws and other characteristics of the legal environment in the state?

Before I summarize the results, certain caveats are in order. The data available do not permit complete answers to these questions. Insurance claims files lack much of the information relevant to the outcome; data that are in principle available are often either missing or reported with error. After I eliminated cases lacking data on key variables, sample sizes are small, especially for cases tried to verdict. The estimation techniques used here do not control for non-random selection of cases by stage of disposition, so the estimates are to be interpreted as conditional on the stage of disposition, not unconditional estimates of population parameters.

Nevertheless, the results are generally consistent with the legal standard and the economic model. Awards are strongly related to the severity of the injury and, for permanent injuries, to the life expectancy of the plaintiff. Consistent with the law of compensable damages, cases involving death receive substantially less than cases of permanent disability. Awards are positively related to earnings for persons employed at the time of the injury. On average, employment status and sex do not have a significant effect, consistent with the legal standard of compensation for potential wages for persons not employed in the market sector.

I do not attempt to measure the adequacy of compensation. I do test the more limited hypothesis that awards increase at least in proportion to pecuniary loss (as would be expected if awards attempt to compensate for both pecuniary and non-pecuniary loss, and the latter is at least proportional to the former). The hypothesis of proportionality cannot be rejected. This implies that the result reported by others, that awards tend to be greater than loss for cases involving relatively low loss and less than loss for cases involving relatively large loss, could be due entirely to error in reported loss.

Awards depart from the legal standard of compensable damages in ways predicted by the economic model of disposition, which predicts that settlements discount expected court awards by the probability that the plaintiff would win. Proxies for the probability that the plaintiff would win (insurer's evaluation of liability, allegation of *res ipsa,* representation by an attorney) are all associated with higher settlements.[1]

As predicted, settlements are also positively related to the litigation costs of the defense: the elasticity of award with respect to number of hospital defendants is .2; for physician defendants it is .4.[2] Settlements are negatively related to the litigation costs of the plaintiff, as represented by the lag in reporting the claim, for cases settled at suit.[3] Awards are positively related to the limits of the defendants' liability insurance coverage, after number of defendants is controlled for.

Measurement of the determinants of the probability that the plaintiff will win were generally less successful than size of awards. However, what little information there is on the evidence of negligence is consistent with the legal standard. The single most powerful predictor of payment being made to the plaintiff, at verdict and in settlement, is the insurance company's evaluation that the claim had legal merit. Cases involving obvious error (1976) or alleging *res ipsa* (1974) have a higher probability of winning, as do cases where the injury was "induced by treatment." Cases involving less obvious error--misdiagnosis or lack of informed consent, incomplete cure or unauthorized treatment--are less likely to receive payment. Representation by an attorney increases the probability of payment.[1]

---

[1] The positive association between attorney representation and settlements probably reflects imperfectly measured loss and positive marginal product of the attorney as well as a higher probability of winning.

[2] The higher value for physicians is attributed to higher time and embarrassment costs of going to court.

[3] This coefficient may overestimate the effect of plaintiff costs because of correlation with imperfectly measured loss and evidence of negligence.

The model predicts that the probability of payment to the plaintiff depends not only on the evidence of negligence but also on the size of the expected award and the litigation costs of the defense and (negatively) on the litigation costs of the plaintiff. Consistent with these predictions, the evidence shows that the probability of payment to the plaintiff out of court is positively related to the severity and duration of the injury and to reported loss (proxy for expected award) and negatively related to the lag in reporting the claim (proxy for plaintiff litigation costs). However, the effect of the number of defendants (proxy for defense litigation costs) is significantly positive only in 1974. Because roughly 50 percent of all claims receive no payment, the allegation that insurance companies can be forced to pay on all claims, no matter how "frivolous," is clearly false.

The model predicts that the propensity to litigate is positively related to the expected award, negatively related to the costs of litigation, and positively related to uncertainty as to the outcome. The data available explain less of the decision to go to verdict than the decision to file suit. The size of the compensable damages has the expected positive effect. Strong evidence of negligence tends to encourage settlement. Cases involving multiple defendants are more likely to go to suit. Thus it appears that the tendency of multiple defendants to raise costs for the defense and thereby encourage settlement is more than offset by the increase in uncertainty and possibly lower costs for the plaintiff.

---

[1] For potential bias in this variable as an indicator of liability, see fn 1, p. 61. However, the measure of the effect of an attorney is downward biased because it is conditional on closing at pre-suit. The higher probability of winning at suit than at pre-suit is in part due to attorney representation, which is virtually 100 percent at suit.

The attempt to measure the effect of state characteristics was generally unsuccessful. These include changes in common law doctrine, such as modification of the locality rule; tort reforms enacted since 1975, such as ceiling on awards and on contingent fees; rulings applying to all branches of tort law, such as adoption of comparative negligence, measurement of pain and suffering by a mathematical formula, differences in wrongful death statutes; per capita income and the average income of lawyers in the state; and the average delay in the state courts. Individual coefficients were sensitive to the specification and unstable across years. This may be due to correlation with other measured and unmeasured variables, imprecise measurement, simultaneous determination, or simply no effect. In the case of the 1975-76 changes in malpractice law, which were enacted after many of the claims used here had been filed, it could not be hoped to measure full, long-run effects. Even as a group, state characteristics added little explanatory power. However, the strong positive effects on the propensity to litigate of the number of claims in the state, the excess limits ratio, and the average plaintiff win rate suggest that there are distinct differences across states not captured by the variables used in this analysis.

Because of the quality of the data, policy conclusions are tentative. Perhaps the most important finding is that the disposition of claims conforms to some extent to the legal standard and departs from it in ways that are consistent with a simple model of rational behavior by the parties involved, subject to the constraints implied by the expected outcome in court and the costs of litigating. Therefore, although the great majority of cases are settled out of court, they are influenced by the legal standards adopted by the courts. The most extreme criticisms, that the system is random with respect to whether and how much payment is made, are clearly unfounded. Evaluation of the tort reforms that have already been enacted requires further analysis of claims closed since 1976.

The finding that cases involving obvious medical error are less likely to be litigated is relevant to the proposal to designate specific injuries that are more likely than not due to negligence as compensable on a no-fault basis. Because the cases likely to be designated as "compensable events" are those less likely to be litigated under the negligence system, the savings in litigation costs may be small. The most costly cases, in terms of litigation, will be left in the negligence system.

Reforms designed to reduce the cost of litigation may be counterproductive because of perverse effects on the incentive to settle. Although the *maximum* the parties would be willing to spend might be reduced, the actual amount spent might not be because the incentives to settle and, hence, spend less than the potential maximum are inversely related to the cost of litigating. A necessary condition for settlement is that disagreement as to the outcome is less than the costs of continuing. This implies that the greatest hope for a reduction in expenditure on litigation lies in reforms that reduce uncertainty as to the standard of care, scope of liability and size of awards, or increase the information available to each side as to the strategy of the other.

Appendix

MODEL OF THE DISPOSITION OF CLAIMS

The model may be summarized as follows:

$$V = V[Z_1; L^p(Z_1, Z_2, w^p); L^d(Z_1, Z_2, w^d)] \qquad \text{(A.1)}$$

= verdict in court.

$$q = W[Z_2; L^p(Z_1, Z_2, w^p); L^d(Z_1, Z_2, w^d)] \qquad \text{(A.2)}$$

= probability of a verdict for the plaintiff.

$$\hat{L}^p = L^p(Z_1, Z_2, w^p) \qquad \text{(A.3)}$$

= plaintiff's optimum expenditure on litigation.

$$\hat{L}^d = L^d(Z_1, Z_2, w^d) \qquad \text{(A.4)}$$

= defendant's optimum expenditure on litigation.

$$A^p = \hat{r}^p(Z_2, \hat{L}^p)\hat{B}^p(Z_1, \hat{L}^p) - w^p\hat{L}^p(Z_1, Z_2, w^p) \qquad \text{(A.5)}$$

= plaintiff's minimum ask.

$$A^d = \hat{r}^d(Z_2, \hat{L}^d)\hat{B}^d(Z_1, \hat{L}^d) + w^d\hat{L}^d(Z_1, Z_2, w^d) \qquad \text{(A.6)}$$

= defendant's maximum offer.

$$F = F(A^p - A^d) \qquad \text{(A.7)}$$

= propensity to litigate.

$$S = gA^d + (1 - g)A^p \qquad \text{(A.8)}$$

= settlement.

Changes in the exogenous variables, Z_1, Z_2, w^p, and w^d affect the outcome variables both directly and indirectly through their effect on litigation inputs, L^p and L^d. Consider first the effect on litigation inputs.

The plaintiff's optimum litigation input, $\hat{L}^p$, is the solution to the equation:

$$G = r^p_L B^p + r^p B^p_L - w^p = 0.$$

For any parameter θ, $\hat{L}^p_\theta \gtreqless 0$ as $G_\theta \gtreqless 0$. Let Z_1 be any variable that tends to increase the size of award, such as severity of the injury.

$$G_{Z_1} = r^p_L B^p_{Z_1} + r^p B^p_{LZ_1} > 0 \text{ and } \hat{L}^p_{Z_1} > 0, \text{ assuming } B^p_{Z_1}, B^p_{LZ_1} > 0.$$

Let Z_2 be any variable that tends to increase the probability of a verdict for the plaintiff, such as convincing evidence of negligence.

$$G_{Z_2} = r^p_{LZ_2} B^p + r^p_{Z_2} B^p_L \gtreqless 0 \text{ and } \hat{L}^p_{Z_2} \gtreqless 0, \text{ assuming } r^p_{Z_2} > 0,\ r^p_{LZ_2} \gtreqless 0.$$

$\hat{L}^p_{Z_2}$ is positive unless an increase in Z_2 reduces the marginal product of L^p in raising r^p, in which case the sign is uncertain.

Let w^p be any variable that raises the plaintiff's cost of inputs, such as the opportunity cost of time of the plaintiff or attorney.

$$G_{w^p} = -1 \text{ and } \hat{L}^p_{w^p} < 0.$$

The defendant's optimum litigation input, $\hat{L}^d$, is the solution to the equation:

$$H = r^d_L B^d + r^d B^d_L + w^d = 0.$$

For any parameter θ, $\hat{L}^d_\theta \lesseqgtr 0$ as $H_\theta \lesseqgtr 0$.

Let Z_1 and Z_2 be as defined previously; w^d is any variable that increases the defendant's cost of inputs, such as opportunity cost of the defendant's time. Then:

$H_{Z_1} = r^d_L B^d_{Z_1} + r^d B^d_{LZ_1} < 0$ and $\hat{L}^d_{Z_1} > 0$, assuming $B^d_{Z_1} > 0$, $B^d_{LZ_1} < 0$.

$H_{Z_2} = r_{LZ_2} B^d + r^d_{Z_2} B^d_L \lesseqgtr 0$ and $\hat{L}^d_{Z_2} \gtrless 0$, assuming $r^d_{Z_2} > 0$, $\gamma^d_{LZ_2} \gtrless 0$.

$\hat{L}^d_{Z_2}$ is negative unless an increase in Z_2 reduces the marginal product of $\hat{L}^d$ in reducing r^d, in which case the sign is uncertain.

$$H_{w^d} = 1 \text{ and } \hat{L}_{w^d} < 0.$$

The effect of changes in the parameters on the minimum ask and maximum offer are as follows:

$$\begin{aligned}
A^p_{Z_1} &= \hat{r}^p[\hat{B}^p_{Z_1} + \hat{B}^p_L \hat{L}^p_{Z_1}] + \hat{B}^p \hat{r}^p_L \hat{L}^p_{Z_1} - w^p \hat{L}^p_{Z_1} \\
&= \hat{r}^p \hat{B}^p_{Z_1} > 0.
\end{aligned}$$

$$\begin{aligned}
A^d_{Z_1} &= \hat{r}^d[\hat{B}^d_{Z_1} + \hat{B}^d_L \hat{L}^d_{Z_1}] + \hat{B}^d \hat{r}^d_L \hat{L}^d_{Z_1} + w^d \hat{L}^d_{Z_1} \\
&= \hat{r}^d \hat{B}^d_{Z_1} > 0.
\end{aligned}$$

$$\begin{aligned}
A^p_{Z_2} &= \hat{B}^p[\hat{r}^p_{Z_2} + \hat{r}^p_L \hat{L}^p_{Z_2}] + \hat{r}^p \hat{B}^p_L \hat{L}^p_{Z_2} - w^p \hat{L}^p_{Z_2} \\
&= \hat{B}^p \hat{r}^p_{Z_2} > 0.
\end{aligned}$$

$$\begin{aligned}
A^d_{Z_2} &= \hat{B}^d[\hat{r}^d_{Z_2} + \hat{r}^d_L \hat{L}^d_{Z_2}] + \hat{r}^d \hat{B}^d_L \hat{L}^d_{Z_2} + w^d \hat{L}^d_{Z_2} \\
&= \hat{B}^d \hat{r}^d_{Z_2} > 0.
\end{aligned}$$

Thus an increase in Z_1 or Z_2 increases the ask and offer by the direct increase in the expected payout. The indirect effect on expected payout, due to an increase in $\hat{L}$, is exactly offset by the incremental expenditure on litigation, $w\hat{L}$.

Similarly, the effect of a change in per unit input costs is confined to the direct effect:

$$A^p_{w^p} = [\hat{r}^p \hat{B}^p_L + \hat{r}^p_L \hat{B}^p - w^p]\hat{L}^p_{w^p} - \hat{L}^p$$

$$= -\hat{L}^p < 0.$$

$$A^d_{w^d} = [\hat{r}^d \hat{B}^d_L + \hat{r}^d_L \hat{B}^d + w^d]\hat{L}^d_{w^d} + \hat{L}^d$$

$$= \hat{L}^d > 0.$$

An increase in input costs tends to reduce the ask and raise the offer.

Finally, the net effect of changes in the parameters on the outcome variables of interest can be derived as follows:

AWARD

$$dV/dZ_1 = \underset{+}{V_{Z_1}} + \underset{+}{V_{L^p}}\,\underset{+}{L^p_{Z_1}} + \underset{-}{V_{L^d}}\,\underset{+}{L^d_{Z_1}} + \underset{+}{V_{L^p_{Z_1}}}\,\underset{+}{L^p} + \underset{-}{V_{L^d_{Z_1}}}\,\underset{+}{L^d} \overset{?}{>} 0,$$

which is positive unless the negative effect of increased defense expenditure offsets the positive direct effect and positive indirect effect through increased plaintiff expenditure.

$$dV/dw^p = V_{L^p}\, L^p_{w^p} > 0.$$

$$dV/dw^d = V_{L^d}\, L^d_{w^d} < 0.$$

PROBABILITY OF A VERDICT FOR THE PLAINTIFF

$$dq/dZ_2 = \underset{+}{q_{Z_2}} + \underset{+}{q_{L^p}}\,\underset{?}{L^p_{Z_2}} + \underset{-}{q_{L^d}}\,\underset{?}{L^d_{Z_2}} \overset{?}{>} 0.$$

$$dq/dw^p = q_{L^p} L^p_{w^p} < 0.$$

$$dq/dw^d = q_{L^d} L^d_{w^d} > 0.$$

PROPENSITY TO DROP THE CASE

The propensity to drop, D, depends on the probability that the ask becomes negative and no positive offer is made. The expected effects are:

$$D_{Z_1} < 0; \; D_{Z_2} < 0; \; D_{w^p} > 0; \; D_{w^d} < 0.$$

PROPENSITY TO LITIGATE

The effect on the propensity to litigate, F, is the effect on the spread between the ask and offer:

$$F_{Z_1} = A^p_{Z_1} - A^d_{Z_1}$$

$$= \hat{r}^p \hat{B}_{Z_1} - \hat{r}^d \hat{B}^d_{Z_1}$$

$$= (r^p - r^d)V_{Z_1} + r^p u^p_{1,Z_1} - r^d u^d_{1,Z_1} \gtrless 0$$

where u_1 = error in predicting V.

$$F_{Z_2} = A^p_{Z_2} - A^d_{Z_2}$$

$$= (B^p - B^d)q_{Z_2} + B^p u^p_{Z_2} - B^d u^d_{Z_2} \gtrless 0$$

where u_2 = error in predicting q.

If u_1 and u_2 have mean zero, but variance proportional to V and q respectively, then F_{Z_1} and $F_{Z_2} > 0$.

$$F_{w^p} = A^p_{w^p} < 0.$$

$$F_{w^d} = A^d_{w^d} < 0.$$

SIZE OF SETTLEMENT

$$S_{Z_1} = gA^d_{Z_1} + (1 - g)A^p_{Z_1} > 0.$$

$$S_{Z_2} = gA^d_{Z_2} + (1 - g)A^p_{Z_2} > 0.$$

$$S_{w^p} = (1 - g)A^p_{w^p} < 0.$$

$$S_{w^d} = gA^d_{w^d} > 0.$$

BIBLIOGRAPHY

AIA, *American Insurance Association, Special Malpractice Review: 1974 Closed Claim Survey,* Insurance Services Office, New York, 1976.

Brown, John P., "Toward an Economic Theory of Liability," *Journal of Legal Studies,* Vol. 2 (1973, pp. 323-350.

Dietz, Stephen, Bruce Baird, and Lawrence Berul, "The Medical Malpractice Legal System," in Appendix to the *Report of the Secretary's Commission on Medical Malpractice,* DHEW No. (OS), pp. 73-89, Washington, D.C., 1973.

Gould, John P., "The Economics of Legal Conflicts," *Journal of Legal Studies,* Vol. 2, 1973, pp. 279-290.

Halperin, Max, William C. Blackwelder, and Joel I. Verter, "Estimation of the Multivariate Logistic Risk Function: A Comparison of the Discriminant Function and Maximum Likelihood Approaches," *Journal of Chronic Diseases,* Vol. 24 (1971), pp. 125-128.

Insurance Services Office (ISO), *Compilation of General Liability Experience: Professional Liability,* New York, 1975.

-----, *1976 Product Liability Closed Claims Survey,* New York, December 1976.

The Institute of Judicial Administration, *Calendar Status Study* (1970, 1972, 1974), 40 Washington Square South, New York, New York, 10012.

Landes, William M., "An Economics Analysis of the Courts," *Journal of Law and Economics,* Vol. 14, 1971.

Munch, Patricia, *Contingent Fees for Personal Injury Litigation,* R-2458-HCFA, The Rand Corporation, April 1980.

-----, "The Costs and Benefits of the Tort System if Viewed as a Compensation System," P-5921, The Rand Corporation, June 1977a.

-----, "An Economic Analysis of Eminent Domain," *Journal of Political Economy,* June 1976.

-----, "Issues in Professional Liability: Multiple Defendants," Unpublished Rand research, 1977b.

-----, *Determinants of the Frequency of Malpractice Claims,* R-2661-HCFA/ICJ (forthcoming).

National Association of Insurance Commissioners (NAIC), *Malpractice Claims,* Vol. 1, No. 4, May 1977.

Posner, Richard A., "An Economic Approach to Legal Procedure and Judicial Administration," 2, *Journal of Legal Studies,* 1973.

-----, "A Theory of Negligence," *Journal of Legal Studies,* Vol. 1, No. 1, 1972.

Schwartz, Murray L., and Daniel J. B. Mitchell, "An Economic Analysis of the Contingent Fee in Personal Injury Litigation," *Stanford Law Review,* Vol. 22, 1970.

Steves, Myron, and A. McWhorter, Jr., "Notes on the Malpractice Insurance Market," *CPCU Annals,* Vol. 28, No. 4, December 1975.

U.S. Department of Transportation, *Automobile Insurance and Compensation Study,* Washington, D.C., 1971.

Westat, Inc., *Medical Malpractice Closed Claim Study, 1976,* Rockville, Maryland, 1978.

-----, *Study of Medical Malpractice Claims Closed in 1970,* Rockville, Maryland, 1973.

CONTINGENT FEES FOR PERSONAL INJURY LITIGATION

PREPARED FOR THE HEALTH CARE FINANCING ADMINISTRATION,
U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

PATRICIA MUNCH DANZON

PREFACE

The research reported here was funded by the Health Care Financing Administration, U.S. Department of Health, Education, and Welfare, under Contract No. 600-76-0150, "An Economic Analysis of Medical Malpractice." Other Rand work performed under this contract includes a study of the disposition of malpractice claims (R-2622-HCFA), an analysis of the determinants of the frequency of malpractice claims (R-2661-HCFA/ICJ), and an examination of the level of malpractice insurance premiums (R-2623-HCFA).

Contingent fees are the typical form of payment for plaintiff attorneys on personal injury cases. Several states have imposed restrictions on contingent fees as one among several measures enacted in response to the growing cost of malpractice claims. This report presents a theoretical analysis of contingent and hourly wage contracts. The analysis is used to show the effect of these different forms of contract on the amount of litigation and to identify the probable effects of restrictions on contingent fees.

SUMMARY

Contingent fees are the typical mode of payment for plaintiff attorneys in personal injury litigation. Under a contingent contract, the attorney's fee is contingent on the outcome of the case: If the plaintiff wins, the attorney is paid some predetermined fraction of the award; if the plaintiff loses, the attorney receives no fee. Under an hourly wage contract, the plaintiff pays the attorney an hourly rate for his time, regardless of the outcome of the case. Thus, the contingent fee shifts the risk of zero return on the investment in litigation from the plaintiff to the attorney.

There is a long tradition of hostility toward contingent fees. A number of states have recently restricted their use in medical malpractice cases, in response to the rise in frequency and severity of malpractice claims. Opponents of contingent fees claim that giving lawyers the right to finance litigation makes them stir up cases, hoping for excessive financial reward. They also claim that the lawyer's financial stake creates a conflict of interest between lawyer and plaintiff. Proponents argue that by shifting the risk to the lawyer, contingent fees give legal recourse to plaintiffs who have legitimate claims but are unwilling to risk investing in an hourly wage contract. Further, they contend that by making the fee depend on the amount of recovery, contingent contracts align the interests of attorney and client. This alignment is particularly important when the client lacks the expertise to assess the lawyer's performance.

There is no conclusive theoretical analysis and little empirical evidence to support claims on either side of the debate. Yet, to the extent that fee arrangements affect the number of outcomes of suits filed, they affect the efficiency of the tort system in the performance of its dual functions of deterring negligent behavior and compensating people injured through negligence.

The purposes of this study are to clarify the debate over contingent fees and to determine the probable effects of restrictions on contingent fees on the frequency and severity of claims. The study analyzes the incentives created by the alternative assignment of risk

under the hourly and contingent contracts. The analysis assumes that attitudes toward risk affect the willingness of plaintiffs and attorneys to engage in litigation with uncertain payoff. Plaintiffs and attorneys may be averse to risk, or may prefer risk, or may be neutral to risk. A risk-averse person is one who would not participate in a lottery if the odds of winning were only even. A risk-preferring person would be willing to participate in a 50-50 lottery. A risk-neutral person is indifferent to the 50-50 lottery. In the medical malpractice context, a risk-averse plaintiff would not bring suit if the expected cost of litigation were as great as the expected award (i.e., the possibility of winning times the amount of the award).

The contractual assignment of risk and the parties' risk preference affect the number of cases that will be filed, the attorney's effort per case, and hence the probability of winning and size of award. If plaintiffs and attorneys are risk-neutral, the frequency of claims, the average gross recovery, and the plaintiff's probability of winning will be unaffected by the type of contract: The amount of litigation will be that necessary to maximize the net recovery of plaintiffs. The division of gross recovery into the net to the plaintiff and fee to the attorney is also unaffected, on average, over all plaintiffs. However, on cases won the actual fee is higher, hence net recovery is lower with a contingent fee than with an hourly wage contract. Obviously, on cases lost, the fee and therefore the plaintiff's out-of-pocket loss are zero.

Under the more plausible assumption that plaintiffs are typically risk-averse and lawyers risk-neutral, the frequency of claims, average gross recovery, and plaintiff's probability of winning will be lower with an hourly wage than with a contingent fee. This is because with an hourly rate, the plaintiff finances the litigation and bears the risk of zero return. With an hourly contract, the risk-averse plaintiff will spend less than is necessary to maximize the net value of his claim.

The analysis disputes the allegation that contingent fees result in excessive (above competitive) rewards for attorneys. Rational allocation of time by an attorney between contingent fee and hourly rate cases and market competition both act to control fees. True, the actual fee on cases won will exceed the value of time spent on those cases. However, the excess is compensation for the ex ante risk of receiving no fee. What little empirical evidence is available confirms that, averaging over cases won and lost, the effective hourly earnings of attorneys paid on a contingent basis are similar to the hourly earnings of defense attorneys paid by the hour. This is consistent with competition and elimination of risk aversion from the contingent fee sector. Competition will tend to drive attorneys who are risk-averse out of the contingent fee sector. The risk-averse attorney would require an expected fee that exceeds the value of his time, whereas a risk-preferring attorney would accept a fee that is less than the value of his time. Thus, risk aversion on the part of the attorney tends to reduce the net recovery of the client, and risk preference tends to increase it.

Evidence on the ratio of legal costs to average award suggests that returns to attorney effort diminish more rapidly for the defense than for the plaintiff in severe injury cases. One implication of this asymmetry in the "judicial production function" is that a reduction in attorney wage rates will tend to induce a greater increase in the efforts of plaintiffs' attorneys and, hence, lead to an increase in the size of awards and plaintiff win rates. This factor may have contributed to the increase in frequency and severity of claims and pro-plaintiff shift in the law in recent years.

Constraints on the minimum time necessary to process a case may bar cases with low expected recovery from suit. This effect is independent of the type of fee contract: If small cases are barred from the judicial process, it is because of the high fixed costs, not the contingent fee system. Ceilings on the contingent fee percentage may significantly reduce the number of hours an attorney will spend on a case and effectively bar certain cases from trial. The types of cases likely to be affected are (1) cases of severe injury but uncertain evidence of negligence, (2) cases with low expected payoff (minor injuries), and (3) cases where the plaintiff is highly risk-averse. All these factors contribute to a relatively large fee percentage, in the absence of constraints.

Restriction on contingent fees would also tend to be regressive, deterring low- and middle-income plaintiffs from filing even meritorious suits. Risk aversion, low wealth, and high borrowing costs all create obstacles to filing that a contingent fee reduces but may not totally eliminate. For example, a person with average risk aversion but assets of less than $20,000 would be unwilling to hire an attorney on an hourly basis to handle a malpractice suit yielding the average award of $29,456 (1977 data), even with a high probability of winning. Thus, in the absence of the contingent fee, the number of cases filed would certainly be less. Conversely, the common allegation that the contingent fee induces attorneys to bring claims with little legal merit has no basis in logic. The fact that the fee depends on winning provides an incentive to screen out cases with little legal merit--an incentive that is lacking with an hourly fee.

If ceilings or outright prohibition of contingent fees are intended to reduce the amount paid out through the malpractice liability system, they will probably be successful. However, if the socially optimal level of litigation is that which would be undertaken by a fully informed, risk-neutral plaintiff, the unconstrained contingent fee is likely to yield the closest approximation to this ideal. Without contingent fees, plaintiff risk aversion will produce a suboptimal investment in litigation and hence suboptimal deterrence and compensation. Even if the explicit policy objective were to reduce frequency of suits, size of awards, and expenditure on litigation, it is doubtful that limiting contingent fees is an efficient means of achieving these results.

ACKNOWLEDGMENTS

I would like to thank Bridger Mitchell, Joseph P. Newhouse, and Charles E. Phelps for very helpful comments on a preliminary version of this report.

CONTENTS

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS ix

TABLES xiii

Section
I. INTRODUCTION 1

II. HOURLY WAGE AND CONTINGENT FEE CONTRACTS WITH RISK NEUTRALITY 5
  Model A: Hourly Wage Contract with Risk Neutrality .. 6
  Model B: Schwartz and Mitchell Contingent Fee Contract with Risk Neutrality 6
  Model C: Reinterpretation of Contingent Fee Contract with Risk Neutrality 9
  Conclusion 11

III. HOURLY WAGE AND CONTINGENT FEE CONTRACTS WITH RISK AVERSION 13
  Model D: Hourly Wage Contract with Plaintiff Risk Aversion 13
  Model E: Contingent Fee Contract with Risk Aversion 15
  Conclusion 18

IV. THE ATTORNEY'S SHARE: INTERPRETATION, EVIDENCE, AND EFFECTS OF CONSTRAINTS 22
  Determinants of the Attorney's Share 22
  Evidence of the Attorney's Share 25
  Ceilings on Contingent Fees 29
  Restrictions on Outright Sale 31

V. INFERENCES FROM FEES TO THE "JUDICIAL PRODUCTION FUNCTION" 33
  Plaintiff 33
  Defense 33

VI. NUMBER OF SUITS 38
  Plaintiff Risk Aversion 38
  Attorney Attitude toward Risk 40

VII. CONCLUSION 41

BIBLIOGRAPHY 45

TABLES

1. Weights Attached to $N_{A,L}$ and $N_{p,L}$ Due to Risk Aversion, by Type of Contract 19

2. Percent of Lawyers Who Normally Use Different Types of Fee Arrangements 26

3. Contingent Fee Percentage Rates 26

4. Sliding Contingent Fee Arrangements 27

5. Disposition of Malpractice Claims 35

I. INTRODUCTION

Contingent fees are the predominant form of payment for plaintiff attorneys on medical malpractice and all other types of personal injury litigation. With a contingent fee, the attorney finances the litigation and is reimbursed for his services only if he wins the case.[†] Thus, the contingent fee effectively shifts from the plaintiff to the attorney the risk of receiving zero return on the investment in litigation.

There is a long tradition of hostility toward contingent fees. They are prohibited in England and Canada. During 1975-1976, 18 states passed legislation limiting contingent fees for medical malpractice cases.[‡] McKinnon (1964) summarizes the basic objections to contingent fees: (1) Giving the lawyer the right to finance litigation allegedly motivates him to stir up suits--suits that are supportable but that would not be brought on the client's initiative and suits that are groundless but have a nuisance value; (2) the stimulus to stir up suits is said to be fueled by excessive financial rewards; (3) the lawyer's stake in the suit creates a conflict of interest between the lawyer and his client, which impedes settlement. The lawyer is allegedly more prone to gamble.

Conversely, proponents argue that (1) the contingent fee provides access to the courts for plaintiffs who are deterred by the cost and risk of hiring an attorney on an hourly basis but who may have legitimate claims; (2) the contingent fee permits an efficient allocation of the risk and cost of investing in litigation with an uncertain outcome; (3) by making the fee depend on the size of the recovery, the contingent contract aligns the interests of the attorney with those of his client. This is particularly valuable when the client lacks the expertise to monitor the attorney's performance.

[†] Contingent fee contracts may stipulate that the plaintiff pay certain fixed costs, such as filing fees, regardless of the outcome. In practice, the attorney often advances these expenses and reimbursement may be de facto contingent on the plaintiff winning.

[‡] Limiting contingent fees has been one of several legislative enactments designed to control the growth in the number of malpractice claims and the size of awards, which culminated in increases in malpractice insurance premiums of over 300% in some states in 1975.

Schwartz and Mitchell (1970) dispute the view that the incentives implied by a contingent fee contract induce the attorney to act in his client's best interest. They compare the incentives of an attorney paid on a contingent basis with a situation where a fully informed, risk-neutral client hires an attorney on an hourly basis. Relative to this benchmark, they conclude that the contingent fee results in fewer hours per case and consequently lower gross recoveries and lower lawyer fees.[†] Their conclusion that the contingent fee induces the attorney to invest fewer hours than would maximize the net value of the claim to the plaintiff is accepted by Clermont and Currivan (1978) and Reder (1978). Reder advocates the abolition of legal restrictions on the outright sale of the claim by the client to the attorney or, as a second best solution, the public provision of a certifier, to provide the plaintiff with disinterested advice on when to settle.

Very little empirical evidence exists on the effects of contingent fees. Using data from a survey of plaintiff and defense attorneys handling medical malpractice cases, Dietz, Baird, and Berul (1973) find that the effective hourly wage of plaintiff attorneys paid on a contingent basis is not much higher than the hourly wage of the defense paid by the hour.[‡] This rough evidence that contingent fees do not affect lawyers' earnings is reinforced by the opinions expressed by attorney respondents to the survey. Less than 20% of either plaintiff or defense lawyers thought that the contingent fee stimulates claims, affects settlements, or results in excessive fees.

Thus, there is no consensus on how contingent fees affect, if at all, the number of claims filed, the propensity to litigate, and the size of awards and settlements. Yet the answers to these questions have important implications for the efficient operation of the tort system. The tort system is designed to perform two primary functions: deterring negligent behavior and compensating persons injured by negligent behavior. To the extent that fee arrangements for plaintiff attorneys affect the number and outcome of suits filed, they affect the efficiency of the system in performing its social purpose of deterrence and compensation. Public policy with respect to contingent fees should therefore be evaluated in the light of its effects on the efficient functioning of the entire tort system.[†]

[†] The model underlying these conclusions is discussed more fully in Section II.

[‡] The effective hourly wage resulting from contingent fee contracts is computed by averaging fees on cases won over hours spent on cases won and lost and adjusting for fees to referring attorneys.

This study develops a theoretical analysis of hourly wage and contingent fee contracts in order to infer the effects of limits on contingent fees. Specifically, the analysis examines the effect of hourly and contingent fee contracts on the amount of attorney hours, given the facts of a case; the probability of the plaintiff winning and the size of the gross recovery; the division of the gross recovery between the net recovery to the plaintiff and the fee of the attorney; and the plaintiff's propensity to bring a claim, given the occurrence of an injury.

Section II presents a theoretical analysis of the effects of hourly wage and contingent fee contracts, assuming risk neutrality. Section III modifies the analysis to allow for nonneutral attitudes toward risk for both plaintiff and attorney. Section IV analyzes the determinants of the attorney's share in a contingent fee contract, presents some empirical evidence, and derives implications about the effects of limits on the fee. Section V applies theory and evidence to draw conclusions about the effectiveness of plaintiff and defense attorney input in influencing court outcomes. Section VI examines the effect of hourly and contingent fees on the number of suits filed from a given universe of injuries. The final section summarizes the results and derives policy conclusions.

[†] If information and transactions are costless, the efficient functioning of the tort system requires that the injurer compensate the injured for damages caused in all cases of injury due to negligence. This zero-transactions-cost ideal is used here as a benchmark against which to evaluate outcomes under alternative fee arrangements. In a world of costly information and transactions, this may not be the appropriate benchmark, but revising the optimality conditions to take account of positive information and transactions costs is beyond the scope of this report.

II. HOURLY WAGE AND CONTINGENT FEE CONTRACTS WITH RISK NEUTRALITY

This section presents models of the hourly wage contract (Model A) and two alternative formulations of the contingent fee contract: the Schwartz and Mitchell formulation (Model B) and a reinterpretation (Model C) based on an alternative specification of the nature of competition for cases. At this point, the analysis takes as given the plaintiff's decision to hire an attorney. It focuses on the effect of the type of contract on the number of attorney hours spent on a case, the implied size of gross recovery, and the division of this gross recovery between the net recovery of the client and the attorney's fee.[†] Risk neutrality of both client and attorney is assumed throughout.

The following notation is used:

p = probability of a positive recovery by the plaintiff.
A = conditional expected gross recovery, given a positive recovery.
w = wage rate of the attorney on an hourly wage contract, which is equal to the opportunity cost of attorney time on a contingent fee contract.
L = input of attorney hours.
α = attorney's share of the award under a contingent fee contract.
$pA(L)$ = payoff function.

By assumption, the expected gross recovery and the probability of a positive recovery are positive functions of attorney input, L, with diminishing returns. Thus,

$$\frac{\partial A}{\partial L}, \frac{\partial p}{\partial L} > 0; \qquad \frac{\partial^2 A}{\partial L^2}, \frac{\partial^2 p}{\partial L^2} < 0.$$

[†] "Recovery" and "award" are used interchangeably to refer to court awards and out-of-court settlements.

MODEL A: HOURLY WAGE CONTRACT WITH RISK NEUTRALITY[†]

The plaintiff is assumed to have full information in that he knows as much as the attorney does about the payoff function, pA(L), and can monitor attorney performance without incurring costs. The plaintiff buys attorney time by the hour.

Decision Rule: The plaintiff selects L to maximize his expected net recovery:

$$\max_{L} \phi = pA - wL. \tag{A1}$$

Maximization of Eq. A1 with respect to L yields the first-order condition:

$$\frac{\partial(pA)}{\partial L} = w. \tag{A2}$$

Equation A2 implies that the informed client maximizes his expected net recovery by purchasing attorney time up to the point where the attorney's gross marginal product, in terms of increased expected recovery, is equal to his hourly wage rate.

MODEL B: SCHWARTZ AND MITCHELL CONTINGENT FEE CONTRACT WITH RISK NEUTRALITY

It is assumed that the plaintiff is ignorant of the payoff function, pA(L), and hence is unable to evaluate the attorney's performance. The attorney is therefore free to choose the hours spent on the case. He is assumed to maximize his expected net profit, given that his fee is some fraction, α, of the gross recovery, and his time has an opportunity cost, w, equal to the wage he could earn on other cases.

Decision Rule: The attorney selects L to maximize his expected net profit:

$$\max_{L} \phi = \alpha pA - wL. \tag{B1}$$

[†] Model A corresponds to the hourly wage contract modeled by Schwartz and Mitchell (1970), except that their assumption of perfect knowledge implies p = 1 and A(L) is a certain, not an expected, value. With risk neutrality, the implications of the models are identical.

Maximizing Eq. B1 with respect to L gives

$$\alpha \frac{\partial(pA)}{\partial L} = w. \tag{B2}$$

Equation B2 implies that the attorney spends time up to the point where *his share* of the gross marginal product is equal to the opportunity cost of his time. Comparing Eqs. A2 and B2, since $\alpha < 1$, equilibrium occurs at a larger gross marginal product with a contingent fee than with an hourly wage contract. Assuming diminishing marginal productivity, this implies fewer hours per case and lower expected recovery with the contingent fee contract.

The Schwartz and Mitchell (S&M) formulation of the contingent fee contract implies that the attorney input is lower, and hence the gross recovery is lower than in the case where a fully informed client buys time by the hour. The S&M conclusions with respect to the effect of the contingent fee on the *net* recovery of the plaintiff and the fee of the attorney depend on a series of assumptions about the market for legal services. S&M assume that attorneys compete for contingent fee cases by bidding down α until hourly earnings on contingent fee cases are reduced to the hourly wage rate in other cases.[†] Taken alone, this assumption implies that both the net recovery of the plaintiff and the fee of the attorney will be lower with a contingent fee than if a fully informed plaintiff hires an attorney by the hour.

However, S&M modify their conclusion that both net recovery and fee will be lower with a contingent fee by introducing a general equilibrium adjustment in the market for legal services. They assume that this market consists of a fixed number of personal injury cases and a fixed number of other cases. They also assume that the number of attorney hours supplied to the market is fixed. Initially, all cases are handled with an hourly wage, and plaintiffs are assumed fully informed. S&M then drop the assumption of informed plaintiffs and introduce the contingent fee for personal injury cases. The analysis of the contingent fee implies that the number of hours spent on personal injury cases decreases. S&M assume that these hours released from personal injury cases because of the introduction of the contingent fee are "dumped" in the hourly wage market, and depress the hourly wage by an amount that depends on the elasticity of demand for hours. This implies a reduction in the opportunity cost of time spent on contingent fee cases and thus some increase in time input, relative to the situation before the general equilibrium reduction in wage rates. Taking into account this second-round effect, S&M conclude that the introduction of the contingent fee unambiguously reduces the attorney's fee on personal injury cases (since both w and L fall). The effect on the plaintiff's net recovery is ambiguous, however, depending on whether the reduction in gross recovery (due to lower L) is greater or less than the reduction in the lawyer's fee. However, if the reduction in w is negligible, because demand for attorney time is elastic, the initial conclusion holds--the net recovery of the plaintiff is lower with a contingent fee.

[†] A reasonable assumption, in view of the mobility of lawyers between contingent fee and hourly wage business.

The S&M conclusion, that the contingent fee contract fails to maximize the net value of the plaintiff's case because of a suboptimal input of attorney time, is analogous to the traditional view that share contracts in general are inefficient because they induce a suboptimal input of factors that are paid on a share basis. Cheung (1968) has shown that this criticism of share contracts in agriculture is misplaced because it ignores competition among tenants and assumes unrealistically simple forms of contract. Cheung shows that actual share cropping contracts in fact stipulate the level of inputs of the parties to the contract, in addition to their shares of output. When inputs are mutually determined, the private incentive for each party to invest less than the mutually optimal level of its input can be mitigated.

In the case of contingent fees for legal services, however, detailed specification of inputs is apparently not common. The typical contract stipulates only the attorney's share, which may vary depending on whether the case goes to trial, and perhaps that the client should assume certain expenses, such as court costs. Precise stipulation ex ante of the attorney's time input is presumably not feasible because the plaintiff faces high costs of ascertaining the optimum input and high costs of monitoring performance. It is shown in Model C below, however, that competition for cases and monitoring of outcomes can provide a perfect substitute for fully informed and costless monitoring of inputs.

MODEL C: REINTERPRETATION OF CONTINGENT FEE CONTRACT WITH RISK NEUTRALITY

All assumptions of the S&M model are retained with one exception. Whereas S&M assume attorneys compete for cases only with respect to the fee percentage, α, here the more plausible assumption is made that attorneys compete in the net recovery they offer a client, $pA(1 - \alpha)$. Thus, bidding is in terms of the expected gross recovery as well as the fee percentage. Each attorney bases his bid on the number of hours he expects to spend. To maximize his chances of getting the case, he will bid a pA and an α that maximize the expected net recovery to the plaintiff, subject to the condition that he cover the opportunity cost of his time. Although the bid is with respect to expected outcome, this implies a commitment of L. The hourly wage rate implied by the expected fee, αpA, must equal what could be earned selling those hours on an hourly basis.

Decision Rule: The attorney selects L and α to maximize the expected net recovery of the client, subject to the constraint that the attorney cover the opportunity cost of his time.[†]

$$\max_{L,\alpha} \phi = pA(1 - \alpha) + \mu(\alpha pA - wL), \qquad \text{(C1)}$$

where μ is a Lagrange multiplier.

Maximizing with respect to L, α, and μ, respectively, gives

$$(1 - \alpha)\frac{\partial(pA)}{\partial L} + \mu\left[\alpha\frac{\partial(pA)}{\partial L} - w\right] = 0, \qquad \text{(C2)}$$

$$-pA + \mu pA = 0, \qquad \text{(C3)}$$

[†] The opportunity cost of attorney time, w, may include some rents.

$$\alpha pA - wL = 0. \tag{C4}$$

From Eq. C3, $\mu = 1$. Substituting this result in Eq. C2 yields

$$\frac{\partial(pA)}{\partial L} = w. \tag{C2'}$$

Equation C2' defines the attorney's optimum input of hours, $\hat{L}$, which in turn determines the expected award, $\hat{p}\hat{A}$, that the attorney will bid to try to get the case. The optimum L is that which equates the opportunity cost of his time, w, with the gross marginal product. This is identical with the condition of equilibrium under an hourly wage contract, when a perfectly informed client determines the number of hours he will buy to maximize the net value of his claim (Eq. A2). Thus, the number of hours and expected gross recovery are invariant to the type of contract.

Although L and pA are invariant with respect to the type of contract, the division of the gross recovery into attorney's fee and net recovery of the client is affected. In comparing the attorney's fee under an hourly wage and a contingent fee contract, the possibility of a zero recovery creates a distinction between the fee expected ex ante and the fee realized ex post on cases with a positive recovery and hence positive fee. Equation C4 shows that the *expected* fee, αpA, is equal to the fee that would have been paid under an hourly wage contract, since w and L are identical. However, the *actual* fee on cases won, αA, exceeds the cost of time by a multiple that is the inverse of the ex ante optimized probability of winning, $\hat{p}$. On medical malpractice cases with attorney representation, roughly 60% of the claims obtain some positive recovery.[†] Assuming this is correctly anticipated, contingent fees on average will appear to overcompensate attorneys 66% for their time spent on cases won.[‡] Thus, if the probability of winning is less than unity, the competitively determined fee will exceed the value of time spent ex post because it was uncertain ex ante. This holds even under the assumption that the attorney is risk-neutral.

[†] Munch (1980).

[‡] $(.6)^{-1} = 1.66$. Reder (1978) states that fees on cases won must compensate the attorney for his time spent on cases lost. This is a misleading interpretation of the determination of α. In a competitive environment, an attorney cannot charge a client for hours spent on other

CONCLUSION

If all parties are risk-neutral, the input of attorney effort and hence the probability of winning and size of gross recovery will be the same with a contingent fee contract (with the plaintiff unable to monitor attorney input) as if a fully informed plaintiff hired an attorney on an hourly basis. The plaintiff's expected net recovery on any particular case and the average net recovery of plaintiffs in general are invariant to the type of contract and are the maximum that can be obtained, taking into account the facts of the case, the payoff to attorney input, and the opportunity cost of attorney time. Similarly, the attorney's expected fee on a particular case and the average fee over all cases are the same with a contingent and an hourly contract. However, the actual fee realized on cases won is higher with a contingent fee by a multiple that is the inverse of the optimized probability of winning. Similarly, the plaintiff's realized net recovery on cases won is lower with a contingent fee. Obviously, the fee is less and the plaintiff's out-of-pocket cost is less on cases lost with a contingent fee.

These conclusions differ from those reached by S&M because of different assumptions about the nature of competition. Both the S&M model and the reinterpretation used here assume that attorneys compete for cases, such that hourly earnings on contingent fee cases are bid down to the hourly wage rate that can be earned on other cases. S&M assume that competition is confined to the fee percentage. Given a competitively determined fee percentage and opportunity cost of his time, the attorney is assumed to determine his time input on a case to maximize his own profit. It follows logically from these assumptions that attorney input with a contingent fee will be less, and hence the gross recovery, net recovery, and attorney's fee will be less with a contingent

---

cases he has lost. Each client will only contract to pay for the expected value of time spent on *his* case.

fee than if a fully informed client hired an attorney on an hourly basis.[†]

The alternative model developed here makes the more plausible assumption that attorneys compete in terms of the expected net recovery offered the client, which depends on the expected gross recovery as well as the attorney's fee percentage. This assumption yields the conclusions listed above, that is, no difference, by type of contract, in attorney effort, gross recovery, expected net recovery and expected fee, and therefore maximization of the net value of the case to the plaintiff. This conclusion of identical outcomes with contingent and hourly contracts holds even if the market for legal services is not competitive and rents are earned by some or all attorneys, provided that the degree of monopoly power is the same for a given attorney, whether he sells his time on an hourly or a contingent basis. With a uniform degree of monopoly power in the hourly wage and contingent fee markets, attorney input will be less, and the gross and net recovery will be lower, than with a competitive market for legal services. However, the outcome will still be the same, regardless of the type of contract.

[†] These conclusions ignore second-order effects discussed above, which, under plausible assumptions about the market for legal services, are too small to alter the conclusions.

III. HOURLY WAGE AND CONTINGENT FEE CONTRACTS WITH RISK AVERSION

In this section, hourly wage and contingent fee contracts are compared under the assumption that the plaintiff and the attorney are averse to risk. Expected utility replaces wealth as the objective to be maximized. The following additional notation is used:

$U(\ )$ = plaintiff's utility function.
V = plaintiff's initial wealth.
$Z(\ )$ = attorney's utility function.
Y = attorney's initial wealth.
Subscript 1 denotes the state in which the plaintiff wins.
Subscript 0 denotes the state in which the plaintiff loses.
Primes denote marginal utilities.
$\bar{U}' = pU_1' + (1 - p)U_0'$ = plaintiff's expected marginal utility.
Z_c = attorney's utility of his certain income alternative, $Y_c = Y + wL$.
$\lambda = -(U''/U')$ = absolute risk-aversion parameter of plaintiff's utility function defined at V.
$\lambda^* = \lambda V$ = relative risk-aversion parameter of plaintiff's utility function defined at V.
$\gamma = -(Z''/Z')$ = absolute risk-aversion parameter of attorney's utility function defined at Y_c.
$\gamma^* = \gamma Y_c$ = relative risk-aversion parameter of attorney's utility function defined at Y_c.
N_{AL} = elasticity of A with respect to L.
N_{pL} = elasticity of p with respect to L.

MODEL D: HOURLY WAGE CONTRACT WITH PLAINTIFF RISK AVERSION

Decision Rule: The plaintiff selects L to maximize his expected utility over the two possible outcomes, a positive or a zero recovery.

$$\max_{L} \phi = pU[V + A - wL] + (1 - p)U[V - wL]. \qquad \text{(D1)}$$

Maximization of Eq. D1 with respect to L yields.

$$pU_1' \frac{\partial A}{\partial L} + \frac{\partial p}{\partial L} [U_1 - U_0] = w[pU_1' + (1 - p)U_0'] . \qquad \text{(D2)}$$

Equation D2 shows that when the plaintiff finances the litigation, risk aversion affects the evaluation of both the expected payoff and the expected costs. To determine the net effect of risk aversion on the plaintiff's purchase of attorney time, rewrite Eq. D2 as

$$p \frac{\partial A}{\partial L} \left(\frac{U_1'}{\bar{U}'} \right) + \frac{\partial p}{\partial L} \left(\frac{U_1 - U_0}{\bar{U}'} \right) = w, \qquad \text{(D2')}$$

where $\bar{U}' = pU_1' + (1 - p)U_0'$ = expected marginal utility.

Using a second-order Taylor expansion of the utility function around U(V) and multiplying through by L/pA,

$$\frac{\partial A}{\partial L} \frac{L}{A} \left[\frac{1 - \lambda(A - wL)}{1 - \lambda(pA - wL)} \right] + \frac{\partial p}{\partial L} \frac{L}{p} \left[\frac{1 - \lambda((A/2) - wL)}{1 - \lambda(pA - wL)} \right] = \frac{wL}{pA} , \qquad \text{(D2")}$$

or

$$N_{A,L} \left[\frac{V - \lambda^*(A - wL)}{V - \lambda^*(pA - wL)} \right] + N_{p,L} \left[\frac{V - \lambda^*(.5A - wL)}{V - \lambda(pA - wL)} \right] = \frac{wL}{pA} ,$$

or

$$N_{A,L}k_1 + N_{p,L}k_2 = \frac{wL}{pA} ,$$

where k_1 = the term in the first square brackets,

k_2 = the term in the second square brackets.

With risk neutrality, $\lambda = 0$ and $k_1 = k_2 = 1$. With risk aversion, $\lambda > 0$, $k_1 < 1$, and $k_2 \gtrless 1$ as $p \gtrless .5$. Thus, the risk-averse plaintiff always discounts $N_{A,L}$, the elasticity of award with respect to attorney input. Hence he will buy fewer hours for purposes of increasing the award and will obtain a lower gross and net recovery than if he is risk-neutral. The reduction is greater, the lower p and the lower V relative to the expected net recovery. Similarly, if $\hat{p}$, the optimized probability of winning, is less than .5, the risk-averse plaintiff discounts the elasticity of p with respect to attorney input. He therefore buys fewer hours for increasing p and is less likely to win than if he is risk-neutral. Again, the effect is greater, the lower p and the lower V relative to the expected net recovery. However if $\hat{p} > .5$, risk aversion increases the plaintiff's evaluation of a further increase in p, and he spends more for this purpose and is more likely to win than if he is risk-neutral.

MODEL E: CONTINGENT FEE CONTRACT WITH RISK AVERSION

Because the contingent fee contract assigns to the attorney the risk of zero return on the investment in litigation, the attitudes toward risk of both the plaintiff and the attorney are relevant to choice of attorney effort and therefore the outcome. Competition implies that an attorney seeks to maximize the expected utility of the plaintiff, subject to the constraint that the attorney's expected utility of the contingent fee is equal to the utility of the alternative, certain income.

Decision Rule:

$$\begin{aligned}\max_{L,\alpha,\mu} &= pU[V + A(1 - \alpha)] + (1 - p)U[V] \\ &\quad + \mu[pZ(Y + \alpha A) + (1 - p)Z(Y) - Z(Y + wL)] \qquad \text{(E1)}\end{aligned}$$

The first-order conditions with respect to L, α, and μ, respectively, are

$$pU_1'(1 - \alpha)\frac{\partial A}{\partial L} + (U_1 - U_0)\frac{\partial p}{\partial L} + \mu\left[pZ_1'\alpha\frac{\partial A}{\partial L} + (Z_1 - Z_0)\frac{\partial p}{\partial L} - Z_c'w\right] = 0, \qquad \text{(E2)}$$

$$-pAU_1' + pA\mu Z_1' = 0, \qquad \text{(E3)}$$

$$pZ(Y + \alpha A) + (1 - p)Z(Y) = Z(Y + wL). \qquad \text{(E4)}$$

From Eq. E3,

$$U_1' = \mu Z_1' .$$

Substituting in Eq. E2 yields

$$p \frac{\partial A}{\partial L} + \frac{\partial p}{\partial L} \left[\frac{U_1 - U_0}{U_1'} + \frac{Z_1 - Z_0}{Z_1'} \right] = w \frac{Z_c'}{Z_1'} . \qquad \text{(E2')}$$

Expanding Eq. E2' by a second-order Taylor expansion of U around V and Z around the certain income alternative, Z_c, and multiplying through by L/pA yields

$$N_{A,L} + N_{p,L} \left\{ (1 - \alpha) \left[\frac{V - .5\lambda^* A(1 - \alpha)}{V - \lambda^* A(1 - \alpha)} \right] + \alpha \left[\frac{Y_c - .5\gamma^*(\alpha A - wL)}{Y_c - \gamma^*(\alpha A - wL)} \right] \right\}$$

$$= \frac{wL}{pA} \left[\frac{Y_c}{Y_c - \gamma^*(\alpha A - wL)} \right] \qquad \text{(E2'')}$$

or

$$N_{A,L} + N_{p,L}[(1 - \alpha)k_3 + \alpha k_4] = \frac{wL}{pA} k_5 ,$$

where k_3 = the term in the first square brackets,
k_4 = the term in the second square brackets,
k_5 = the term in the third square brackets.

A comparison of Model E and Model C shows the effects of plaintiff risk aversion with a contingent fee contract. To isolate the effect of plaintiff risk aversion, assume that the attorney is risk-neutral, that is, $k_4 = k_5 = 1$. If the plaintiff is averse to risk, k_3 exceeds unity. Thus, with a contingent fee contract, the risk-averse plaintiff values an increment in the size of award as if he were risk-neutral but places a higher value on an increment in p than if he were risk-neutral. The effect of plaintiff risk aversion with a contingent fee is to increase L beyond $\hat{L}$, the optimal level with risk neutrality (with either a contingent fee or an hourly wage contract), but the additional hours are devoted exclusively to increasing the probability of winning. The gross recovery will be the same, and the probability of winning will be greater than in the risk-neutral case.

However, the net recovery of the plaintiff, both expected and realized, is reduced by his risk aversion which leads him to purchase hours in excess of the risk-neutral optimum, $\hat{L}$. This is because the marginal product of hours in excess of $\hat{L}$ is less than their cost, since $\hat{L}$ is defined as the point at which $\partial(pA)/\partial L = w$. The reduction in net recovery may be viewed as the price that the risk-averse plaintiff pays for the "insurance" of increasing the probability of winning. This "insurance" is paid by an increase in the attorney's share of the award. This follows from the equilibrium condition (Eq. E4) which, with attorney risk neutrality, is simply

$$\alpha = \frac{wL}{pA} .$$

Since $N_{pA,L} < 1$ at $L = \hat{L}$, it follows that α must increase if $L > \hat{L}$.

A comparison of Model D and Model E shows that, in contrast with the risk-neutral plaintiff, the risk-averse plaintiff is affected by the type of contract. If the plaintiff is risk-averse, the gross recovery is greater with a contingent fee. The probability of winning on claims with $\hat{p} < .5$ (presumably when the plaintiff's case is relatively weak) is higher with a contingent fee. On claims with $\hat{p} > .5$, the comparison is ambiguous. The risk-averse plaintiff's *expected* net recovery (and the average for all plaintiffs) is higher with a contingent fee. It is possible that the *realized* net recovery on cases won is also higher. This contrasts with the risk-neutral case, where it was shown that although the *expected* net recovery is invariant with respect to the type of contract, the *realized* net recovery on cases won is lower with a contingent fee than with an hourly fee. With a risk-averse plaintiff, realized net recovery is greater with a contingent fee as

$$A_{HW} - wL_{HW} < A_{CF}(1 - \alpha) = A_{CF} - \frac{wL_{CF}}{p} ,$$

where the subscript HW denotes hourly wage,

the subscript CF denotes contingent fee.

Since $A_{HW} - wL_{HW} < A_{CF} - wL_{CF}$, this condition holds for values of p close to 1.

Thus far the analysis has assumed that the plaintiff is risk-averse and the attorney risk-neutral. To isolate the effects of attorney risk aversion ($\gamma^* > 0$), assume that the plaintiff is risk-neutral ($\lambda^* = 0$). If the attorney is risk-averse, γ^* is positive and k_4 and k_5 exceed unity. The risk-averse attorney assigns a higher value to increments in p and a higher cost to time spent on a risky project. Multiplying through Eq. E2' by k_5^{-1} yields

$$N_{A,L}k_5^{-1} + N_{p,L}[(1 - \alpha)k_5^{-1} + \alpha k_4 k_5^{-1}] = \frac{wL}{pA} . \qquad \text{(E2''')}$$

Since $0 < k_5^{-1} < k_4 k_5^{-1} < 1$, attorney risk aversion induces discounting of both components of the payoff, $N_{A,L}$ and $N_{p,L}$, and hence there is an investment of less time on increasing both A and p. This implies that awards and plaintiff win rates are lower if attorneys using contingent fee contracts are risk-averse rather than risk-neutral. Because the net recovery of the plaintiff is maximized with risk neutrality, it follows that the plaintiff is worse off with a risk-averse attorney.[†] Table 1 summarizes the effects of plaintiff and attorney risk aversion on the weights attached to $N_{A,L}$ and $N_{p,L}$ under the hourly wage and contingent fee contracts. Weights less than one indicate a decrease in attorney time input relative to the risk-neutral case; weights greater than one indicate an increase in attorney time input.

CONCLUSION

If the plaintiff is risk-averse, the outcome is not invariant with respect to the type of contract, in contrast with the risk-neutral case. When the plaintiff is risk-averse and the attorney is risk-neutral, the

---

[†] If the plaintiff is also risk-averse, his preferences partially offset the attorney's tendency to discount $N_{p,L}$ and may dominate if λ^* is sufficiently larger relative to γ^*.

Table 1

WEIGHTS ATTACHED TO $N_{A,L}$ AND $N_{p,L}$ DUE TO RISK AVERSION, BY TYPE OF CONTRACT

	Hourly Wage	Contingent Fee	
Elasticity of Payoff	Plaintiff Risk-Averse $\lambda^* > 0$	Plaintiff Risk-Averse, Attorney Risk-Neutral $\lambda^* > 0,\ \gamma = 0$	Plaintiff Risk-Averse, Attorney Risk-Averse $\lambda^* > 0,\ \gamma^* > 0$
$N_{A,L}$	$\frac{V - \lambda^*(A - wL)}{V - \lambda^*(pA - wL)} < 1$	1	$\frac{Y_c - \gamma^*(\alpha A - wL)}{Y_c} < 1$
$N_{p,L}$	$\frac{V - \lambda^*(.5A - wL)}{V - \lambda^*(pA - wL)} \gtrless 1$ as $p \gtrless .5$	$\alpha + (1 - \alpha)\frac{V - \lambda^*.5A(1 - \alpha)}{V - \lambda^*A(1 - \alpha)} > 1$	$\Big\{\alpha \frac{[Y_c - .5\gamma^*(\alpha A - wL)]}{Y_c} + (1 - \alpha)$ $\times \left[\frac{Y_c - \gamma^*(\alpha A - wL)}{Y_c}\right] \left[\frac{V - \lambda^*.5A(1 - \alpha)}{V - \lambda^*A(1 - \alpha)}\right]\Big\} \gtrless 1$

NOTES: $\lambda^* = -(U''/U')V$ = plaintiff risk aversion at V, the certain level of wealth.
$\gamma^* = -(Z''/Z')Y_c$ = attorney risk aversion at Y_c, the certain level of wealth.
$N_{A,L}$ = elasticity of A with respect to L.
$N_{p,L}$ = elasticity of p with respect to L.
Weights greater (less) than 1 imply an increase (decrease) in attorney input relative to the risk-neutral optimum, $\hat{L}$.

input of attorney effort to increase the size of the award and hence the gross recovery will be greater with a contingent than with an hourly fee. This is because when the risk-averse plaintiff bears the risk of no return on the investment in litigation, the utility he assigns to an increment in the award is less than its dollar value. The probability of winning will be higher with a contingent fee on claims in which the plaintiff's case is relatively weak ($\hat{p} < .5$). When the plaintiff's case is relatively strong ($\hat{p} > .5$) the risk-averse plaintiff will purchase more attorney effort than if he were risk-neutral with both an hourly and a contingent fee.

With risk aversion, the plaintiff's expected net recovery (or the average recovery for all plaintiffs) is definitely greater with a contingent fee, and the realized net recovery on cases won may even be higher, in contrast with the risk-neutral result where the realized net recovery is lower with a contingent fee. The reason is that whereas with a risk-neutral plaintiff the expected net value of the claim is maximized with either an hourly or a contingent fee, with risk aversion the expected net value of the claim is not maximized, and the distortion is greater with an hourly wage than with a contingent fee.

Comparing the outcomes with a contingent fee under alternative assumptions about attitude toward risk, the expected and realized gross recovery is unaffected by plaintiff risk aversion. However, the risk-averse plaintiff implicitly buys more attorney effort to increase the probability of winning. This "insurance" is paid for in the form of a larger fee percentage for the attorney.

The analysis of the case in which the attorney is nonneutral toward risk contradicts the common notion that a contingent fee contract creates a conflict of interest between a risk-averse plaintiff and an attorney with a greater tolerance of risk. Attorney risk tolerance benefits the plaintiff by inducing greater input and hence a higher expected recovery. Essentially, a risk-tolerant attorney provides the risk-averse plaintiff with insurance in the form of a smaller share of a larger gross recovery. Thus, the contingent fee segment of the market for legal services will tend to attract risk-tolerant attorneys, not necessarily because they earn monetary or nonmonetary rents from gambling but because they offer plaintiffs a higher net recovery than can risk-averse attorneys.[†] Because competition will tend to eliminate risk-averse attorneys from the contingent fee business, attorney risk neutrality is assumed hereafter.

These results hold for the case in which p, the probability of a nonzero recovery, is a continuous, nondecreasing function of attorney effort, L. This is a realistic assumption for the great majority of cases that are closed by settlement, including during trial, rather than by trial to verdict.[‡] The model does not encompass the situation where p is a discontinuous, possibly decreasing function of L. This situation arises when the decision is whether or not to reject a positive settlement offer (p = 1) in order to pursue litigation to court, where p may be considerably less than one, but the expected award, if positive, exceeds the settlement offer.[††]

[†] Alternatively, the benefits of risk neutrality may be obtained by pooling risk through the formation of partnerships or firms.

[‡] Less than 10% of medical malpractice cases are tried to verdict (NAIC, 1977).

[††] In medical malpractice cases tried to verdict, the defendant wins approximately 80% of the time (NAIC, 1977). Data are not available on the percentage of cases litigated to verdict that received a positive settlement offer.

IV. THE ATTORNEY'S SHARE: INTERPRETATION, EVIDENCE, AND EFFECTS OF CONSTRAINTS

This section provides an interpretation of the attorney's fee percentage, α, and shows how it is affected by the characteristics of the case and attitudes toward risk. Empirical evidence on fee percentages is presented. This framework is then used to evaluate the effects of constraints on contingent fees.[†]

DETERMINANTS OF THE ATTORNEY'S SHARE

When the attorney is risk-neutral, his fee is equal to the opportunity cost of his time, adjusted for the ex ante optimized probability of winning (Eq. C4):

$$\alpha A = \frac{w\hat{L}}{p} .$$

If the plaintiff is also risk-neutral (Model C), the wage rate is equal to the expected marginal product (Eq. C2). It follows that α is equal to the elasticity of expected award with respect to attorney effort:

$$\alpha = \frac{\hat{L}}{pA}\frac{\partial(pA)}{\partial L} = \hat{N}_{pA,L} ,$$

where "^" indicates values at the risk-neutral optimum. This is analogous to the solution in Cheung (1968) that the equilibrium rental percentage in a share contract is equal to the elasticity of output with respect to land. Since α is typically one-third, $N_{pA,L} = .3$ is assumed in the following analysis.

The S&M model of the contingent fee equilibrium input, with competitive elimination of rents, implies

$$\alpha \frac{\partial(pA)}{\partial L} = w = \frac{\alpha pA}{L} .$$

Hence $N_{pA,L} = 1$.

[†] The discussion presumes that the expectations on the basis of which α is determined are realized on average.

Since α is typically .3, this model implies that the attorney's gross marginal product is equal to three times his wage rate. In other words, if the plaintiff were to switch to an hourly contract, his expected *net* recovery would be \$2 for every \$1 spent on legal fees. Foregone profits of this magnitude seem implausibly large.

When the plaintiff is risk-averse but the attorney is risk-neutral, α is still equal to the probability-adjusted opportunity cost of the attorney's time. However, the risk-averse plaintiff places a higher value on increasing p and thus purchases greater attorney effort, L*, by offering a larger share, α^*. At the margin, the attorney's wage exceeds his marginal product. Therefore, the fee percentage exceeds the elasticity of output with respect to attorney effort:

$$\alpha^* = \frac{wL^*}{p^*A} > N^*_{pA,L} .$$

Thus, other things being equal, the negotiated α will be higher, the more risk-averse the plaintiff.[†]

To summarize these results, the α that is determined by voluntary, contractual arrangements will be higher: the lower the ex ante optimized probability of winning; the smaller the ex ante optimized award; the higher the opportunity cost of the attorney's time or, equivalently, the greater his productivity; and the more risk-averse the plaintiff.

So far it has been assumed that the attorney's time input is infinitely divisible. In practice there may be some minimum effort required to handle a case. The time required to file briefs, take depositions, negotiate, etc. are to some extent independent of the value of the case. If the minimum time required exceeds the unconstrained optimum, α must rise to compensate the attorney for his additional effort.[‡]

[†] In this case, different "prices" to different plaintiffs does not reflect price discrimination by the attorney. It is a market response to different plaintiffs seeking different products (more effort, surer award) depending on risk aversion.

[‡] The behavior of the defense affects the minimum effort required to ensure the plaintiff a "reasonable" probability of winning. If the defense devotes more time to defending cases involving high stakes, the minimum plaintiff effort may also rise with the stakes. Thus, although it seems more likely that a minimum time constraint will be binding on cases with low stakes, it is possible that cases with higher stakes are also affected.

To illustrate the order of magnitude of the effect on α, consider the case in which both plaintiff and attorney are risk-neutral. If L is constrained to exceed $\hat{L}$ by x%, α must increase if the additional time produces less than an x% increase in the expected recovery, in order to satisfy the constraint that the attorney covers the opportunity cost of his time. The elasticity of α with respect to L is

$$N_{\alpha,L} = \frac{d \ln \alpha}{d \ln L} = 1 - \frac{d \ln pA}{d \ln L} = 1 - N_{pA,L}.$$

If $N_{pA,L} = .33$ at the unconstrained L, $N_{\alpha,L} = .66$. Thus, if the constrained minimum time input is twice the unconstrained optimum, α must increase from .33 to .55. In this example, the attorney would require over half the expected recovery to make it worth his while to take the case.

Thus, if the complexity of the legal process sets a lower bound on the time required to file suit, cases for which this minimum exceeds the optimum may be effectively barred from prosecution because the fee required by an attorney to recover the opportunity cost of his time would absorb too much of the gross recovery and might even exceed it.[†] Any procedural reform that reduces the minimum effort will tend to reduce fees on cases for which the existing constraints are binding and to encourage filing of cases that are currently excluded because either p, the probability of winning, or A, the expected recovery, are too small, in view of the constrained lower bound on L. For example, if the effect of substituting arbitration for the judicial forum for handling cases is to reduce the minimum number of hours needed to bring a case, the model would predict an increase in the filing of cases for minor injuries (low A) and cases with weak evidence for the plaintiff (low p).

It must be emphasized, however, that to the extent that the existence of minimum fixed costs of bringing suit deters filing of some claims, the result derives from the size of the legal fees, relative to the expected gross recovery, and is independent of whether the attorney is paid on an hourly wage or contingent fee basis. It has been demonstrated in Section II that, assuming risk neutrality, the expected gross recovery, the expected attorney's fee, and hence the expected net recovery are the same under an hourly wage or a contingent fee contract. Thus, the contingent fee per se is irrelevant to the exclusion of cases from the system.

[†] To some degree, the constraint can be circumvented by hiring an attorney with a lower opportunity cost of time, w.

EVIDENCE OF THE ATTORNEY'S SHARE

The contingent fee percentage is generally set by a voluntary contract between the plaintiff and the attorney, subject to the influence of custom, recommended fee schedules, and statutory restrictions. McKinnon (1964) states

> A typical contingent fee contract is . . . frequently a form contract containing a customary rate of fee used by lawyers of the community. It may well be based upon a recommendation in a bar association fee schedule, and the client may be unaware of the possibility of negotiating a lower rate.

Although the intent of recommended fee schedules may be to eliminate competition in setting fees, evidence from a survey of contingent fees of medical malpractice cases suggests that there is considerable variation in the fee percentages charged. Dietz, Baird, and Berul (DB&B) (1973) report evidence from two surveys: (1) a National Survey of a random sample of 800 lawyers in private practice, and (2) a random Selective Survey of 400 additional lawyers known to be engaged in medical malpractice litigation.† Table 2 shows the percentage of lawyers who normally use fixed contingent, sliding contingent, and hourly fees; Table 3 shows the distribution of fee percentages for those using a fixed percentage; and Table 4 shows the distribution of sliding fee arrangements.

Approximately 90% of the respondents use some form of contingent fee. Of those using a fixed percentage, $33\frac{1}{3}\%$ is most common, but the

---

† The Selective Survey sample consists of lawyers who specialize in malpractice work.

Table 2

PERCENT OF LAWYERS WHO NORMALLY USE DIFFERENT TYPES OF FEE ARRANGEMENTS

Fee Arrangement	National Survey Respondents[a]	Selective Survey Respondents[b]
Fixed contingent fee	60	45
Sliding contingent fee	30	47
Hourly rate	3	1
Other	7	7
Total	100	100

SOURCE: DB&B (1973).

[a] 30 plaintiff lawyers reporting.

[b] 132 plaintiff lawyers reporting.

Table 3

CONTINGENT FEE PERCENTAGE RATES

Fixed Contingent Fee Percentage	National Survey Respondents (%)[a]	Selective Survey Respondents (%)[b]
25	0	2
$33\frac{1}{3}$	78	58
40	11	15
45	0	5
50	11	20
Total	100	100
Mean fee	36	38
Median fee	36	38

SOURCE: DB&B (1973).

[a] 18 plaintiff lawyers reporting.

[b] 60 plaintiff lawyers reporting.

Table 4

SLIDING CONTINGENT FEE ARRANGEMENTS

Sliding Contingent Fee Arrangement	Frequency of Occurrence	
	National Survey	Selective Survey
20%-25% before trial; 33⅓% through trial	2	1
25% before trial; 33⅓% through trial; 40% through appeals	1	4
33⅓%-50% for adults; 25%-33⅓% for dead or minor plaintiffs		8
33⅓% before trial; 40% through trial and appeal	2	9
33⅓% before trial; 40% through trial and appeal (25%-33⅓% for dead or minor plaintiffs)		7
33⅓% before trial; 50% with trial and appeal	1	5
33⅓% before trial; 40% with trial; 50% through appeal		6
40% through trial; 50% appeal (25% for dead or minor plaintiffs)		3
40% through trial; 50% appeal	1	3
40% before trial; 45%-50% if tried or appealed	1	4
25% before filing; 33⅓% after filing; 40% after trial; 50% on appeal (Oregon State Bar Minimum Fee Schedule)		1
New Jersey Sliding Scale		8
Other		7

SOURCE: DB&B (1973).

range runs from 25% to 50%. Of those using a sliding scale, the most common percentages are $33\frac{1}{3}$% before trial and 40% through trial and appeals, but the range runs from 20% before trial to 50% through appeals.

Thus, these surveys indicate considerable diversity in fee percentages.[†] The pattern of charging a higher percentage for cases taken to trial than for cases settled before trial is consistent with the model, which predicts that the fee percentage is higher, the higher the probability of a zero recovery, because plaintiff win rates are substantially lower on cases tried to verdict than on cases settled out of court.[‡] However, because both the lawyer's time input and, usually, the expected award (if positive) are also greater if the case is tried to verdict, it is not possible to determine the extent to which each of these three factors contributes to the higher percentage charged for cases taken to trial.

These medical malpractice surveys did not ask explicitly whether the percentage charged is related to the size of recovery.[††] It is interesting that a 1961 study of contingent fees for 3000 personal injury cases in New York City shows the same average fee as the medical malpractice surveys (36%) and a distinct inverse relation between fee percentage and size of recovery, both overall and at each stage of disposition. Aggregating over cases settled and tried, the average fee ranges from 41% of cases under \$1000 to 29% of cases over \$25,000.[‡‡]

[†] These surveys show variation in fees nationwide. It is possible that there could be considerable variation among communities, but little variation within a community. State licensure requirements may impede equalizing earnings across states.

[‡] Munch (1980).

[††] DB&B (1973) report that "in many cases the attorneys specified that the normal sliding contingent fees are reduced when the plaintiff is a minor or deceased." Relatively large recoveries could explain the fee reduction in these cases.

[‡‡] Franklin, Chanin, and Mark (1961). In this sample of cases, the fee percentage does not vary systematically by stage of disposition for a given size of recovery. Plaintiff win rates by stage of disposition are not reported.

CEILINGS ON CONTINGENT FEES

The effect of ceilings on contingent fees depends on the extent to which they impose binding constraints. Based on the above evidence on the range of fees typically charged, the Federal Tort Claims Act ceiling of 25% is likely to be binding in most cases. The most stringent limit on fees for medical malpractice cases is the Indiana limit of 15% of any recovery from the patient compensation fund. Limits that decrease with the size of the recovery (sliding scale) are most likely to be binding on severe injury cases. For example, the California medical malpractice statute imposes a ceiling of 40% on the first \$50,000; $33\frac{1}{3}$% on the next \$50,000; 25% on the next \$100,000; and 10% on amounts over \$200,000.[†] Because the average award in California on cases paid in 1976 was \$31,508, the average case at that time would have been unaffected by the ceiling. Unless the limits are revised, however, inflation will bring an increasing number of cases within the binding limits. The California sliding scale yields an average percentage of less than one-third on all cases over \$140,000. For example, the average percentage on an award of \$1 million would be roughly 14%. Assuming that 40% is typical of the freely contracted percentage for a \$1 million case going to trial, the California statute imposes a ceiling that is roughly one-third of the free market fee.

A binding fee ceiling is expected to reduce the number of hours an attorney is willing to spend on a case, and thus reduce both gross and net recovery. A very crude estimate of the order of magnitude involved can be obtained from the model described above. Assuming a risk-neutral attorney, the condition of competitive equilibrium, $\alpha = wL/pA$, can be used to solve for the reduction in L and pA that is necessary to accommodate an imposed reduction in α, given a relation between L and pA.

[†] California State Assembly Bill AB 1, 1976. The total award to which these percentages are applied is the gross award before deduction of the plaintiff's nonrecoverable costs. The California Citizens' Commission on Tort Reform recommended a ceiling of 40% for the first \$50,000; 25% for amounts in excess of \$50,000 if the case goes to trial; and for cases settled out of court the corresponding percentages are 33% and 20%. These percentages are to be applied to net awards, after deducting the plaintiff's nonrecoverable costs (*Righting the Liability Balance*, 1978).

Using the approximation $N_{pA,L} = .33$, the elasticity of pA with respect to α, $N_{pA,\alpha}$, can be obtained as follows:

$$\frac{d \ln pA}{d \ln \alpha} = \frac{d \ln pA}{d \ln L} \frac{d \ln L}{d \ln \alpha} .$$

However,

$$\frac{d \ln \alpha}{d \ln L} = 1 - \frac{d \ln pA}{d \ln L} = .67.$$

Therefore,

$$\frac{d \ln pA}{d \ln \alpha} = .33(.67)^{-1} = .5.$$

Thus, if $N_{pA,L} = .33$ when α is cut by 50%, the equilibrium adjustment of attorney time implies that the expected gross recovery will fall by 25%. If $N_{pA,L} = .4$, then $N_{pA,\alpha} = .66$; that is, the expected gross award falls by two-thirds of the reduction in α, or 33%. These rough calculations imply that binding ceilings on contingent fees will probably induce a considerable reduction in attorney effort and hence gross recoveries. If the constraints are structured as a decreasing percentage of successive increments of award, with no distinction by stage of disposition, cases most likely to be affected are those with large stakes, especially those with a low probability of winning, and especially those that would have been taken to trial. One predicted effect of such constraints is to bar from trial cases of severe injury in which the issue of negligence is in doubt. Positive settlement offers will be made less frequently and for lower amounts.[†] To the extent that such cases serve as precedents for higher awards or extend the scope of liability for negligence, there may be repercussions in the form of lower awards and reduced frequency of claims in general. The ceilings may also be binding on cases in which the unconstrained α is high because the expected recovery is small or because the plaintiff is extremely risk-averse. A decrease in size of awards and frequency of suit is therefore also predicted for minor injuries.

[†] The relation between expected outcome in court and size and likelihood of settlement is analyzed in Munch (1980). Even if few cases are affected by the fee ceilings, there may be a large effect on total pay out because the distribution of payments is highly skewed. For claims closed July 1975 to June 1976, 3% of the claims accounted for 57% of the total indemnity paid to plaintiffs (NAIC, 1977).

RESTRICTIONS ON OUTRIGHT SALE ($\alpha = 1$)

Shavell (1978) and others have shown that if, in a contractual relation between a principal and an agent, the principal is risk-averse and the agent is risk-neutral, the optimal agent's share is one; that is, the principal's utility is maximized by outright sale of his rights to the claim to the agent. In practice, outright sale of personal injury claims does not occur; in fact, based on the evidence from DB&B (1973), α never exceeds .5.

The obstacles to the attorney acquiring more than a 50% share of a claim are both legal and pragmatic. The common law doctrines of maintenance and champerty restrict the attorney's right to acquire a financial interest in a case he is representing. McKinnon (1964) cites evidence that courts would not enforce a fee in excess of 50%.

However, even if these legal restrictions on the sale of claims were abolished, as advocated by Reder (1978), practical obstacles remain. Outright sale is an optimal solution only if the value of the claim depends solely on the behavior of the attorney. In reality, the value of a personal injury claim depends on the behavior of the plaintiff before and during trial. If the plaintiff has no share in the outcome of the claim, he has no incentive to maximize its value. At the limit, the profit-maximizing plaintiff could capture the value of the claim from both the attorney and the defendant by selling his right to the claim to his attorney for its expected value and then selling his services to the defense to save them that expected value. Such a scenario is not totally fanciful. The possibility of profit from changing sides is demonstrated in cases involving multiple defendants, when one defendant settles secretly with the plaintiff and then works for the plaintiff against the other defendants. The secret settlement provides that the final payment by the settling defendant to the plaintiff is inversely related to the award obtained against the remaining defendants.[†]

Thus, the problem of enforcing that the plaintiff acts in a way that maximizes the value of the claim may be as important as the legal restrictions in limiting the share of the claim that the attorney will buy.

[†] Such agreements, called "Mary Carter" agreements, have been outlawed in some states.

V. INFERENCES FROM FEES TO THE "JUDICIAL PRODUCTION FUNCTION"

Assuming optimizing behavior, evidence on legal fees can be used to draw inferences about the productivity of attorney time, pA(L), hereafter referred to as the "judicial production function."

PLAINTIFF

It has been shown that, with risk neutrality, the contingent fee percentage is equal to the optimized elasticity of expected award with respect to attorney effort. Thus, if an attorney charges the same α on all cases of widely different value, it follows that the optimized elasticity of output is uniform. Furthermore, because $\alpha = wL/pA$ and because for each attorney the opportunity cost of time is constant, a uniform α implies that L/pA is also uniform across cases. In other words, if the optimal time input increases in proportion to the expected value of the case, charging a constant α does not imply excessive returns on cases with high stakes, in contrast with the presumption underlying sliding scale limits on contingent fee percentages.

DEFENSE

The defense counsel is usually paid on an hourly basis.[†] The objective of the defense is assumed to be to minimize the expected pay out, which includes both the expected award to the plaintiff and legal fees.[‡] Let H denote defense attorney hours and pA(H) the payoff function with $\partial(pA)/\partial H < 0$, $\partial^2(pA)/\partial H^2 > 0$.

[†] When the defense is an insurance company or a firm, choice of an hourly rather than a contingent fee rate is consistent with the notion that the contingent fee is an efficient device for shifting the risk and cost of financing litigation to the party with the lowest cost and/or most information. When the defendant is a private party, comparable to the plaintiff, the reasons for a different choice of contract are not obvious.

[‡] This formulation ignores the fact that the defendant may be an insurance company that handles many cases. A precedent established on one case may have value for the defense of others, which creates interdependence between the optimum strategies on similar cases.

Decision Rule: The defense selects H to minimize total expected pay out:

$$\underset{H}{\text{Min}} = pA(H) + wH. \tag{F1}$$

The first-order condition for a minimum is

$$\left|\frac{\partial(pA)}{\partial H}\right| = w. \tag{F2}$$

Multiplying through by H/pA yields

$$-N_{pA,H} = \frac{wH}{pA}.$$

Thus, the ratio of legal costs to expected award is equal to the elasticity of expected award with respect to attorney time. This is identical with the equilibrium condition for the plaintiff. A measure of the optimized elasticity of expected award with respect to defense effort can be obtained from the ratio of defense expenditure to award, averaged over cases with zero and nonzero payment to the plaintiff.

The ratio of defense expenditure to award on medical malpractice cases by severity of injury is reported in Table 5. The inverse relation is striking. Defense expenditure decreases from over two-thirds of the award for injuries categorized at Temporary Insignificant to less than 10% of the award for injuries categorized at Permanent Significant. By contrast, the limited evidence on fee percentages on the plaintiff's side shows a narrower range of fees (25%-50%), no systematic inverse relation between fee percentage and expected size of award, and an increase in fee percentage for cases going to trial (which tend to result in relatively large awards).

This evidence suggests an asymmetry in the judicial production function facing the plaintiff and the defense that increases with the severity of the injury. The marginal product of attorney effort appears to diminish more rapidly for the defense than for the plaintiff the more severe the injury. This is not surprising in view of the law defining compensable damages. Economic loss (potential wage loss plus medical

Table 5

DISPOSITION OF MALPRACTICE CLAIMS

Severity of Injury	Claims with Payment (%)	Average Indemnity per Incident ($)[a]	Defense Expense ÷ Indemnity per Incident
Emotional Only	19	821	.91
Temporary Insignificant	31	719	.66
Temporary Minor	39	2,076	.47
Temporary Major	50	7,676	.22
Permanent Minor	52	10,941	.24
Permanent Significant	52	31,701	.12
Permanent Major	47	71,253	.09
Permanent Grave	54	126,418	.08
Death	41	19,081	.15

SOURCE: NAIC (1977).

[a] Includes cases with zero payment.

and other out-of-pocket expenses) sets a floor below which the award cannot fall but leaves the upper limit undefined lacking a clear measure of fair compensation for pain and suffering. It is presumably easier to establish a case for compensation for pain and suffering the more severe the injury.

This evidence on the characteristics of the judicial production functions for plaintiff and defense has important implications for the effect of a change in attorney wage rates on the size of awards and frequency of plaintiff recoveries. If the judicial production functions were entirely symmetrical, a fall in attorney wage rates would induce an equal increase of attorney effort on both sides of a case and result in no net change in outcomes.[†] However, when the judicial production functions are asymmetrical, the effect of an equiproportionate change in wage rates of defense and plaintiff counsel is not neutral in its effect on case outcomes.[‡]

Rough estimates of the effect on outcomes of a change in w may be obtained from the models and data presented so far. For each side, taking the behavior of the other and hence the judicial production function as given, the elasticity of expected award with respect to attorney wage rate is

$$N_{pA,w} = N_{pA,L} N_{\hat{L},w} . \tag{G1}$$

Differentiating the first-order conditions yields the intuitively obvious result that the elasticity of $\hat{L}$ with respect to w is simply the inverse of the elasticity of marginal product with respect to L:

[†] This assumes no change in expectations of the judicial production function. Presumably, over time both sides would realize that a downward shift in the production function had occurred because of the increased effort by the other side. Optimum effort by both sides would therefore fall, possibly to the level prevailing before the fall in wage rates. A complete equilibrium analysis of the effect of a change in attorney wage rates is beyond the scope of this report.

[‡] There is sufficient mobility of lawyers between the defense and plaintiff's bar to justify treating the market for legal services as a single market.

$$N_{\hat{L},w} = N_{MP,L}^{-1} .$$

These results apply equally to the hourly wage or the contingent fee contract, assuming risk neutrality. For a constant elasticity production function of the form

$$|pA| = L^{\delta},$$

the elasticity of marginal product is

$$N_{MP,L} = N_{pA,L} - 1.$$

Substituting in Eq. G1 yields

$$N_{pA,w} = \frac{N_{pA,L}}{N_{pA,L} - 1}$$

The use of .33 and .2 as typical values of $N_{pA,L}$ for the plaintiff and defense, respectively, implies values of $N_{pA,w}$ of .49 for the plaintiff and .25 for the defense. Thus, a fall in attorney wage rates that affects both sides equally will induce a greater increase in attorney effort on the plaintiff's side and hence an increase in awards and in the plaintiff's probability of winning.

Empirical evidence supports this prediction, at least with respect to award size.[†] An analysis of the determinants of the size of awards in a cross-section of medical malpractice claims shows that award size is negatively related to the average income of lawyers in the state, after controlling for severity of injury, economic loss, and other variables.[‡]

[†] The prediction with respect to the plaintiff's probability of winning cannot be tested using average win rates. The prediction relates to the probability of winning for a specified degree of negligence. A decrease in attorney wage rates is likely to attract into the system cases with a lower degree of negligence. Thus, the average plaintiff win rate on all cases could fall, despite an increase in the probability for each case, because of the change in the mix of cases to include more with less evidence of negligence.

[‡] Munch (1980).

VI. NUMBER OF SUITS

PLAINTIFF RISK AVERSION

Section III analyzed the effect of risk aversion, assuming that the plaintiff had decided to hire an attorney, on the amount of attorney time purchased. It was shown that with an hourly wage contract, risk aversion reduces the number of hours purchased. This section examines the effect of plaintiff risk aversion on the decision to hire an attorney with an hourly wage contract. It has been widely recognized that risk aversion may deter a plaintiff from filing a suit when recovery is less than certain if he must pay the attorney on an hourly basis and bear the risk of zero return on his investment.

The risk-averse plaintiff will hire an attorney on an hourly wage basis if the expected utility of the gamble exceeds the utility of his certain wealth.[†] Thus,

$$pU[V + A - wL] + (1 - p)U[V - wL] \geq U[V]. \qquad \text{(H1)}$$

Taking a second-order Taylor expansion around V, Eq. H1 reduces to

$$pA - wL \geq \frac{\lambda^*}{2V} [pA^2 - 2pAwL + (wL)^2]. \qquad \text{(H1')}$$

Dividing both sides by (pA - wL) and substituting a value of 2 for λ* yields the following necessary and sufficient condition for hiring an attorney on an hourly wage basis:[‡]

$$V \geq \left[A - wL + (1 - p) \frac{AwL}{pA - wL} \right]. \qquad \text{(H2)}$$

[†] Certain wealth, V, may be defined to include any payment the plaintiff receives for the claim if he does not hire an attorney. The evidence from the closed-claims surveys shows that an attorney is indispensable to filing a formal suit. The attorney wage cost, wL, incorporates the optimum decision with respect to L.

[‡] For evidence that relative risk aversion is approximately 2, see Brown and Deaton (1972).

Since the third expression in square brackets is positive, a necessary condition for filing suit is that the plaintiff's certain wealth, V, exceed the net recovery expected if the suit prevails, A - wL. The lower p, the greater the excess of V over A - wL that will be necessary to induce the plaintiff to take the gamble.

In 1977 the average award on paid claims was $29,456.[†] This implies an average net recovery of roughly $20,000. Thus, a plaintiff with average risk aversion but with assets of less than $20,000 would be unwilling to hire an attorney on an hourly basis to file a malpractice suit of average size, even with a high probability of winning. The deterrence to filing is greater: the lower the probability of winning; the lower the plaintiff's initial wealth position; the greater the aversion to risk; and the higher w.[‡]

This rough calculation suggests that a substantial number of low- and middle-income plaintiffs would be deterred from filing even "meritorious" claims in the absence of contingent fees, or some equivalent contingent form of payment or legal insurance. The argument is therefore substantially correct: The contingent fee gives those of relatively low means easier access to the judicial process. Restrictions on contingent fees are regressive in their effect.

Even with the contingent fee contract, plaintiff risk aversion may still affect the propensity to sue if filing suit requires a substantial amount of the plaintiff's time, or if the plaintiff has obtained a settlement offer without representation. In deciding whether to hire an attorney, the plaintiff must weigh the expected increase in award against the attorney's fee. This decision can be stated in terms of Eq. H1 by interpreting V to include the settlement offer without an attorney, A as the increment expected to be gained by hiring an attorney, and wL to include any costs incurred by the plaintiff. Thus, even with the contingent fee, plaintiff risk aversion may reduce the number of suits filed below the level that would prevail if plaintiffs were risk-neutral.

[†] NAIC (1977), p. 105.

[‡] High borrowing costs are equivalent to a high w.

ATTORNEY ATTITUDE TOWARD RISK

Probably the most common criticism of the contingent fee is that it encourages attorneys to gamble and stir up "nonmeritorious" or "nuisance" suits. The expected utility hypothesis postulates that an attorney will accept a case if the expected utility of the fee exceeds the expected utility of his alternative income. Therefore, the expected fee must exceed or fall short of the opportunity cost of his time, depending on whether he is averse to or prefers risk. Risk-preferring attorneys will be attracted to the contingent fee business because their risk preference tends to maximize the net recovery of plaintiffs.

Thus, the limited truth in the charge that a gambling attorney stirs up nonmeritorious cases is that a risk-preferring attorney would accept a case that would not be accepted by a risk-neutral attorney, and would charge an expected fee that is less than the cost of his time. Note, however, that a risk-averse attorney's incentive to take nonmeritorious cases is greater with an hourly wage contract, where he bears no risk of zero fee.

A common counterargument is that there is effectively little threat of zero recovery and zero fee and thus little incentive to screen cases, because an insurance company can always be forced to make some payment, to avoid more costly litigation. The evidence presented in Table 5 strongly refutes this allegation. Over 50% of claims are closed without payment to the plaintiff, and, even for severe injuries, the percentage without payment exceeds 40%.[†]

In conclusion, to the extent the contingent fee increases the number of suits, this is due to the reduction of barriers created by plaintiff risk aversion and high financing costs. These barriers are not totally eliminated. Conversely, the attorney has a greater incentive to bring cases with little legal merit if he is paid an hourly fee than if his fee is contingent on winning.

[†] Data for 1970, 1974, and 1976 show a similar plaintiff win rate of roughly 50%.

VII. CONCLUSION

This study examined two alternative forms of contract for purchasing legal services for personal injury plaintiffs: the hourly wage contract and the contingent fee. The contracts differ in their assignment of the risk of zero return on the investment of attorney time. This difference in risk assignment may affect: (1) the decision to bring a case; (2) the amount of attorney effort and hence the probability of winning and the gross recovery on the case; and (3) the division of this gross recovery between the net recovery of the plaintiff and the attorney's fee.

It has been shown that, given certain assumptions about competition in the market for cases, the two forms of contract are identical in outcome if all parties are risk-neutral, with the exception of the distribution of the award. The *expected* fee and *expected* net recovery are the same, with either a contingent fee or an hourly wage contract; however, the *realized* fee on cases won is greater, and hence the *realized* net recovery is lower with a contingent fee.

Under the more reasonable assumption that the plaintiff is risk-averse and the attorney risk-neutral, the input of attorney effort, hence the gross recovery, and the plaintiff's probability of winning will be higher with the contingent fee.[†] The plaintiff's expected net recovery is certainly higher with a contingent fee, and realized net recovery on cases won may even be higher, in contrast with the case of a risk-neutral plaintiff. The reason for this result is obvious: Because investment in attorney effort with an uncertain payoff is risky, less will be undertaken if the risk must be borne by a risk-averse plaintiff, as with an hourly fee.

Risk aversion on the part of the attorney using a contingent fee tends to reduce the expected utility of the plaintiff; risk preference tends to increase it. This is because the risk-averse attorney will require an expected fee that exceeds the opportunity cost of his time as compensation for risk-bearing. By contrast, the risk-preferring attorney will accept an expected fee that is less than the opportunity cost of his time. Consequently, competition will drive out of the contingent fee sector of the market those attorneys who are averse to risk, precisely because they fail to maximize the net recovery of plaintiffs. What limited empirical evidence is available confirms that the effective hourly earnings of plaintiff attorneys, paid on a contingent basis, are not very different from the hourly earnings of defense attorneys paid by the hour. This is consistent with competition and elimination of risk aversion from the contingent fee sector.

[†] Under certain circumstances, the probability of winning may be higher with an hourly wage. See Section II.

The conclusion reached here--that with a risk-neutral plaintiff the outcome is invariant with respect to the type of contract, whereas the risk-averse plaintiff is unambiguously better off with a contingent fee--conflicts with the conclusion reached by Schwartz and Mitchell (1970) and accepted by Clermont and Currivan (1978). The difference in conclusions arises from differing assumptions about the form of competition. These authors assume competitive bidding only on the attorney's share, α.[†] The assumption here is that attorneys compete on the net recovery offered to plaintiffs, which encompasses both the plaintiff's share, $1 - \alpha$, and the magnitude of the expected recovery, pA. If it is accepted that competition on the net recovery to the plaintiff is more meaningful, hence more plausible, it follows that the contingent fee induces the attorney to maximize the net utility to the plaintiff of his claim, subject to the constraint that the attorney cover the opportunity cost of his time. Even if the market for legal services is not perfectly competitive, with the result that attorneys earn more than competitive returns, the above result of invariance with respect to the type of contract still holds, except that the opportunity cost of time includes some rent. This rent would be paid under either type of contract.

The contingent fee reduces, but may not totally eliminate, the obstacles to filing claims created by plaintiff risk aversion, low wealth, and high borrowing costs. It was estimated that a plaintiff with average risk aversion but with assets less than $20,000 would be unwilling to hire an attorney on an hourly basis to handle a malpractice suit yielding the average award of $29,456 (1977 data), even with a high probability of winning. Thus, in the absence of the contingent fee, the number of claims filed would certainly be significantly less. Conversely, the common allegation that the contingent fee induces attorneys to bring claims with little legal merit has no basis in logic. The fact that the fee depends on winning provides an incentive to screen out cases with little legal merit--an incentive that is lacking with an hourly fee.

[†] They also assume competition in hourly wage rates on other business, but this is not fundamental to their conclusion.

For both the plaintiff, using a contingent fee, and the defense, using an hourly wage, the ratio of legal costs to expected award is a measure of the elasticity of expected payoff with respect to attorney effort. Evidence on the ratio of legal costs to expected award suggests that returns to attorney effort diminish more rapidly for the defense than for the plaintiff on severe injury cases. One implication of this asymmetry in the "judicial production function" is that a symmetrical reduction in attorney wage rates will tend to induce a greater increase in attorney effort on the plaintiff's side and hence lead to an increase in size of awards and plaintiff win rates.[†]

It has been shown that constraints on the minimum number of hours required to handle a case may bar cases with low expected recovery from suit, but this effect is independent of the type of contract. Statutory ceilings on the attorney's fee percentage may significantly reduce the number of hours an attorney will spend on a case and effectively bar certain cases from going to trial. The types of case that are likely to be affected are (1) cases involving severe injuries but with relatively low probability of winning; (2) cases where the plaintiff is highly risk-averse; and (3) cases with very low expected payoff. All these factors contribute to a relatively large unconstrained α.

[†] The extent to which this asymmetry underlies the recent pro-plaintiff changes in legal rules defining scope of liability and compensable damages is an interesting question to be addressed in future research.

In conclusion, if the objective of ceilings or outright prohibition of contingent fees is to reduce the amount paid out through the malpractice liability system, some success in achieving this goal is probable. Two major questions remain unanswered, however. First, if the benchmark of the ideal number of cases and optimum expenditure on litigation is that which would be chosen by a fully informed, risk-neutral plaintiff, the unconstrained contingent fee is likely to induce the closest approximation to this ideal. Without the contingent fee, plaintiff risk aversion will produce a suboptimal investment in pursuing claims.[†] Second, even if the explicit policy objective were to reduce the frequency of suits, size of awards, and expenditure on litigation, it is doubtful whether limits on contingent fees are an efficient or equitable means of achieving this result.

[†] The question of whether the benchmark of the risk-neutral plaintiff is the socially correct benchmark, within the context of the overall social objectives of the tort system, is beyond the scope of this report.

BIBLIOGRAPHY

Brown, Alan, and Angus Deaton, "Surveys in Applied Economics: Models of Consumer Behavior," *Economic Journal* 82, December 1972.

California Citizens' Commission on Tort Reform, *Righting the Liability Balance*, Sacramento, Calif., 1978.

Cheung, Stephen N. S., "Private Property Rights & Sharecropping," *Journal of Political Economy*, 1968.

Clermont, Kevin M., and John D. Currivan, "Improving on the Contingent Fee," *Cornell Law Review* 63(4), April 1978.

Dietz, Stephen, Bruce C. Baird, and Lawrence Berul, "The Medical Malpractice Legal System," in Appendix to the *Report of the Secretary's Commission on Medical Malpractice*, Department of Health, Education, and Welfare, Washington, D.C., 1973, pp. 73-89.

Franklin, Marc A., Robert H. Chanin, and Irving Mark, "Accidents, Money and the Law: A Study of the Economics of Personal Injury Litigation," *Columbia Law Review* 61 (1), January 1961.

McKinnon, F. B., *Contingent Fees for Legal Services*, Aldine Publishing Company, Chicago, Ill., 1964.

Munch, Patricia, *The Disposition of Medical Malpractice Cases*, The Rand Corporation, R-2622-HCFA, July 1980.

National Association of Insurance Commissioners (Milwaukee, Wisc.), *Malpractice Claims* 1 (4), May 1977.

Reder, Melvin W., "Contingent Fees in Litigation with Special Reference to Medical Malpractice," in Simon Rottenberg (ed.), *The Economics of Medical Malpractice*, American Enterprise Institute for Public Policy Research, Washington, D.C., 1978.

Schwartz, Murray L., and Daniel J. B. Mitchell, "An Economic Analysis of the Contingent Fee in Personal Injury Litigation," *Stanford Law Review* 22, 1970.

Shavell, Steven, "Risk Sharing & Incentives in the Principal & Agent Relationship," Discussion Paper #619, Harvard University, Cambridge, Mass., May 1978.

*U.S. GOVERNMENT PRINTING OFFICE: 1981-0-720-018/4491

Health Care Financing Grants and Contracts Reports

U.S. Department of Health and Human Services
Richard R. Schweiker, **Secretary**

Health Care Financing Administration
Carolyne K. Davis, **Administrator**

Office of Research, Demonstrations, and Statistics
James M. Kaple, **Acting Director**

Jean LeMasurier, **Director, Program Planning and Support**

Karen Pelham O'Steen, **Research Publications Coordinator**
Donna L. Eskow, **Writer-editor**
Carol J. Pianalto, **Writer-editor**
Alice L. Young, **Writer-editor**
Cynthia Dingle, **Editorial Assistant**

The statements and data contained in this report are solely those of the contractor or grantee and do not express any official opinion of or endorsement by the Health Care Financing Administration.

Send changes of address or requests for this publication to:
ORDS Publications, Rm 1E9 Oak Meadows Building, 6340 Security Blvd., Baltimore, MD 21235.

DEPARTMENT OF
HEALTH AND HUMAN SERVICES
HEALTH CARE FINANCING ADMINISTRATION
BALTIMORE, MARYLAND 21207

OFFICIAL BUSINESS
PENALTY FOR PRIVATE USE. $300

CMS LIBRARY
3 8095 00012207 3

U S Department of
Health and Human Services
Health Care Financing Administration
HHS Publication
HCFA Pub No 03103 May '98